# Metabolic Engineering of Oleaginous Yeast for the Production of Biofuels



by

Mitchell Tai

B.S. Chemical Engineering, Carnegie Mellon University (2006)
M.S. Chemical Engineering Practice, Massachusetts Institute of Technology (2009)

Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemical Engineering
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2012



Signature of author.....
Department of Chemical Engineering
June 22, 2012

Certified by.....
Gregory N. Stephanopoulos
Willard Henry Dow Professor of Chemical Engineering and Biotechnology
Thesis Supervisor

Accepted by.....
Patrick Doyle
Professor of Chemical Engineering
Chairman of the Committee for Graduate Students

# Metabolic Engineering of Oleaginous Yeast for the Production of Biofuels

by

Mitchell Tai

Submitted to the Department of Chemical Engineering on June 22, 2012,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering

## Abstract

The past few years have introduced a flurry of interest over renewable energy sources. Biofuels have gained attention as renewable alternatives to liquid transportation fuels. Microbial platforms for biofuel production have become an attractive option for this purpose, mitigating numerous challenges found in crop-based production. Towards this end, metabolic engineering has established itself as an enabling technology for biofuels development.

In this work we investigate the strategies of metabolic engineering for developing a biodiesel production platform, utilizing the oleaginous yeast *Yarrowia lipolytica* as the host organism. We establish new genetic tools for engineering *Y. lipolytica* beginning with an expression vector utilizing the genetic features from translation elongation factor 1-$\alpha$ (TEF). Additionally, a complementary plasmid was developed allowing for multiple plasmid integration. Bioinformatics analysis of intronic genes in hemiascomycetous yeast also identified relationships between functional pathways and intron enrichment, chronicling the evolutionary journey of yeast species.

Next gene targets were examined within the lipid synthesis pathway: acetyl-coA carboxylase (ACC), delta9-desaturase (D9), ATP citrate lyase (ACL), and diacylglycerol acyltransferase (DGA). A combinatorial investigation revealed the order of contribution to lipid overproduction (from strongest to weakest): DGA, ACC, D9, ACL. Scale-up batch fermentation of selected strains revealed exceptionally high lipid accumulation and yield. These results demonstrate the balance between cellular growth and lipid production which is being modified through these genetic manipulations.

We next explored utilization of alternative substrates to expand the capabilities and utility of *Y. lipolytica*. For xylose, a prevalent substrate in cellulosic feedstocks, expression of the redox pathway from *Scheffersomyces stipitis* and adaptation led to successful substrate utilization. Through the use of cofermentation, growth and productivity on xylose was improved dramatically with xylose-to-lipids conversion successfully demonstrated. For acetate, a potentially useful substrate for electrofuel production, lipid production using our strongest performing strain resulted in high lipid accumulation and yield.

From this study, metabolic engineering of *Y. lipolytica* was successfully used to achieve exceptional lipid overproduction from a variety of substrates. Our genetic tools and recombinant strains establish a strong platform for the study and development of microbial processes for the production of biofuels.

Thesis Supervisor: Gregory Stephanopoulos
Title: Willard Henry Dow Professor of Chemical Engineering and Biotechnology

## Acknowledgments

There are several people to whom I am indebted to for their support during my graduate school career. First, I would like to thank my thesis advisor, Greg Stephanopoulos, for his patience and visionary optimism for developing my enthusiasm for new ideas and new projects. He has allowed me the resources to freely pursue my interests, explore my own ideas, and learn how to be an independent researcher and scientist. This freedom has allowed me to take full ownership and pride in my work. I would also like to thank my thesis committee members - Daniel Wang, Gerry Fink, Kris Prather - for their guidance and feedback on my research and impressing upon me the importance of being systematic and rational in all my scientific investigations.

Next I would like to thank the members of the Stephanopoulos Lab for their help in training, collaborating, and commiserating with me throughout my long academic journey of 12,250 work hours and 360 experiments. I would like to thank Curt Fischer, for his guidance and training in the lab, and most of all for his help in establishing my electronic notebook - something which I am especially proud of. Next I would like to thank Haoran Zhang, who's help in molecular cloning enabled me to break through one of the more frustrating periods of cloning failure - I doubt I would have finished in time without his help. I would also like to thank Andy Silverman and Sagar Chakraborty, who have the advantage of learning from my mistakes and have become enthusiastic successors to my work. I would also like to thank Hang Zhou, who collaborated closely with me on the xylose work, and proved to be an invaluable resource for troubleshooting everything from PCR reactions to broken instruments. Keith Tyo, Benjamin Wang, Jason Walther, Christine Santos, Daniel Klein-Marcuschamer, Adel Ghaderi, Vikram Yadav, Tom Wasylenko - I thank you all for making the lab such an interesting and unique place to be.

I would also like to thank my friends and family for their support throughout this process. I thank my parents, H.T. and Judy, for consistently pushing me to "broaden my horizons" because there's "no free lunch." I thank my brother and sister-in-law, Lawrence and Mazie, for their visits to Boston in order to feed me exceptionally expensive cuisine.

Lastly, I would like to thank Joyce, who has been with me through the most difficult and most joyous times of my graduate career. I am grateful for your energy and enthusiasm for life (an infectious trait which I am lucky to feed off of) and I will always love and respect you for being such an übermensch.

To those that see this:
Thanks for reading my thesis!
It'll be a quick read ;)

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Metabolic Engineering for Biofuel Production

In recent years, concerns over the sustainability and renewability of using fossil fuels for energy have increased interest in alternative fuel sources. As alternatives for transportation, fuels such as biologically-derived ethanol and biodiesel have grown in prominence. However, the demand for large production volumes presents a significant challenge for developing these fuels, as the conversion of substrate to product requires highly optimized yields and productivity. Metabolic engineering provides the unique ability to formulate new solutions to these challenges by optimizing biological systems towards becoming economically viable systems.

Presently, biodiesel is primarily produced by the transesterification of vegetable oils and animal fats. In the United States, it is primarily produced from soybeans. It is estimated that the United States would have to produce 0.53 billion $m^3$ of biodiesel annually to replace the current consumption of all transport fuels. Waste cooking oil, animal fat and even oil crops cannot realistically be produced in quantities of this magnitude. Table 1.1 shows the average oil yield per hectare for various crops. Also estimated is the amount of land area required for crops in order to meet just 50% of the total transport fuel needs at these yields Chisti (2007).

As of 2006, the price of biodiesel derived from palm oil was only $2.50 per gallon Chisti (2007). The corresponding price of petroleum diesel is $3.96 per gallon EIA (2008). While biodiesel is already beginning to be economically competitive, sustainable utilization of land requirements remains an issue. The use of microorganisms for the production of lipids, hydrocarbons, and other complex oils presents an option for producing sufficient quantities of biodiesel. Production of oils from microorganisms is not directly constrained by land usage and surface area. Land requirements for the production of substrate feedstock remain, but these can typically be produced in higher yields or potentially from cellulosic sources.

Table 1.1: Comparison of various plant sources for biodiesel production. Area signifies the requriements for meeting 50 % of all transport fuel needs of the United States as of 2006. Adapted from Chisti 2007.

| Crop | Oil yield (L/ha) | Land Area (M ha) | Percent of existing US Cropping Area (%) |
|---|---|---|---|
| Corn | 172 | 1540 | 846 |
| Soybean | 446 | 594 | 326 |
| Sunflower | 952 | 280 | 153 |
| Canola | 1190 | 223 | 122 |
| Castor | 1413 | 187 | 102 |
| Jatropha | 1892 | 140 | 77 |
| Coconut | 2689 | 99 | 54 |
| Oil palm | 5950 | 45 | 24 |

## 1.2 Oleaginous Yeast as Production Platform

The choice of microbial strain as the host for metabolic engineering is extremely important. One can utilize a model organism with flexible and robust genetic tools or work with a specialized organism evolved for a targeted task, but typically lacking in knowledge or tools for engineering. In both of these approaches the selection of the appropriate organism for the production of oils for biodiesel is an important task. Many factors that must be considered for oil production in microorganisms: genetic manipulability, growth rates, native lipid accumulation, and facility under industrial conditions.

The organism *Yarrowia lipolytica* interestingly presents itself as a robust compromise between these two choices. As an oleaginous yeast, it is among a group of organisms that can naturally accumulate lipids beyond 20% their cell mass. It is one of the most well-studied 'non-conventional' yeast species, with a fully sequenced genome, and an array of genetic tools first created in the 1980's.

The availability of an organism such as *Y. lipolytica* that is naturally capable of elevated lipid production and genetic tools for genetic engineering provides an ideal situation for metabolic engineering to achieve higher lipid production. Metabolic engineering relies heavily

on both the scientific tools available and the organism's metabolic pathways. As such, *Y. lipolytica* represents an attractive candidate and platform for metabolic engineering.

## 1.3 Thesis Objectives and Overview

The main objective of this thesis is to evaluate the use of metabolic engineering to improve the oleaginous yeast *Y. lipolytica* for its ability to heterotrophically convert feedstock carbon substrates into stored triglycerides. Initially, we aim to extend existing genetic tools for genetic manipulation of this organism through development of overexpression platforms and strategies for multiple exogenous gene expression. We then aim to take a rational strain development approach to systematically examine a number of targets for lipid synthesis overproduction. Finally we aim to extend these results into new opportunities by exploring the ulitization of alternative substrates as feedstock.

Chapter 2 opens to provide an overview of tools and strategies of metabolic engineering as well as a survey of the use of metabolic engineering in renewable fuels research. This chapter seeks to provide context for the usefulness and potential of metabolic engineering techniques for this specific field as well as introduce a number of both common issues and strategies pervasive throughout the field.

Chapter 3 focuses on the use of *Y. lipolytica* as the host strain and platform for metabolic engineering. It will discuss the advantages of using such an organism as well as overview the current understanding of the organism's metabolism, physiology, and genetics. Particularly unique or characteristic aspects of the organism are highlighted and will be often compared to the more scientifically ubiquitous organism, *Saccharomyces cerevisiae.*

In Chapter 4 we will overview existing genetic tools for the organism, particularly in relation to gene overexpression. This includes transformation vectors and characterized promoters commonly used for recombinant DNA cloning. It will then discuss the design, preparation and characterization of an expression platform which utilizes endogenous in-

trons for enhanced gene expression. Furthermore, a number of strategies for multiple-gene expression which are implemented later on in the thesis will be introduced and detailed.

Chapter 5 briefly explores the use of bioinformatics to investigate the relationship between introns and gene function. The distribution of expression-enhancing introns in organisms can have several implications towards their evolution and function.

Chapters 6 and 7, detail efforts for the metabolic engineering of the lipid biosynthesis pathway in *Y. lipolytica*. Beginning with identification of rational gene targets within the pathway, Chapter 6 will investigate the effects of two important gene targets on lipid accumulation. Chapter 7 will extend these results by introducing combinatorial expression of these genes and others to investigate any emergent or synergistic effects. Chapter 6 will focus on and highlight an especially productive gene expression combination.

Chapters 8 and 9 will look to extend the lipid production platform away from glucose towards alternative substrates and how these substrates can affect lipid accumulation. Chapter 8 will discuss the construction and characterization of a heterologous xylose utilization pathway. Combining this pathway with elements of the lipid production augmentations from Chapters 6, the performance of these strains are evaluated. Additionally, cofermentation of multiple substrates is examined for its effects on productivity to further explore the potential of these recombinant strains. Chapter 9 details preliminary work utilizing the developed lipid production strain from Chapter 6 for the production of lipids from acetate as an alternative substrate.

Finally we conclude with Chapter 10 to discuss the significance of current work and make recommendations for future work.

# References

Chisti, Y. (2007) Biodiesel from microalgae. *Biotechnology Advances 25*, 294–306.

EIA, Energy Information Administration. U.S. Gasoline and Diesel Fuel Prices. 2008.

# Chapter 2

# Metabolic Engineering: Enabling Technology for Biofuels Production

## 2.1 Engineering the Future of Biofuels

The past few years have introduced a flurry of interest over renewable energy sources. Biofuels have attracted attention as renewable alternatives to liquid transportation fuels. There are numerous potential advantages over fossil fuels: sustainable supply, diversification of energy sources, energy independence and security, rural development, and reduction in greenhouse emissions (Stephanopoulos, 2007). However achieving adequate scale requires a tremendous effort in research and development beyond what has thus far been achieved. The field of metabolic engineering is well-suited to developing the future technologies which will give us widespread, cost-effective, and sustainable transportation fuels. Metabolic engineering is the improvement of cellular activities by manipulation of metabolic networks through the use of recombinant DNA technology (Bailey, 1991). Interdisciplinary advances in metabolic engineering have yielded powerful strategies and methods to understand and manipulate whole metabolic pathways with confidence (Stephanopoulos, 1999; Alper and Stephanopoulos, 2007).

To date, numerous efforts have successfully engineered and optimized metabolic networks to produce high-value targets for use in the pharmaceutical and fine chemicals industries (Keasling, 2010). However, attention is now being turned towards commodity-scale processes which require both cost-efficiency and robustness (Stephanopoulos, 2007). Currently, the most prevalent biofuels are ethanol produced from corn or sugarcane, and biodiesel produced from vegetable oils. Under current production processes, however, both biofuels struggle to compete economically and find difficulty integrating into existing petroleum-based technologies and infrastructure (Hill et al., 2006). Two developmental challenges underpin these shortcomings: the need for a better feedstock, and the need for a better fuel. However these challenges also represent key opportunities to develop the next generation of biofuel technologies. A central element in these technologies will be the use of metabolic engineering to develop the biological platforms which produce these biofuels.

### 2.1.1 Engineering for Improved Feedstocks

For the past few years, production of ethanol from corn and biodiesel from vegetable oils has been increasing rapidly. Last year, US production capacity of corn ethanol exceeded 13.0 billion gallons per year (bgy), approaching 10% of the national gasoline demand. Meanwhile, global biodiesel production is approaching 5.0 bgy, with a majority coming from Europe. However, production of these biofuels from plants like corn or rapeseed also competes for arable crop land needed for food. This adds undesirable price sensitivities between biofuels and food, and has already shown adverse effects on food prices. Transforming forests or existing crop land can also sometimes have the effect of increasing greenhouse gas emissions, counteracting the carbon emissions benefit of biofuels (Fargione et al., 2008). The primary operating cost for producing biofuels is the cost of the feedstock: 60% in the case of corn ethanol and 80% for soybean biodiesel (Pimentel and Patzek, 2005; Balat and Balat, 2008). Even with gains in process yield, current crop-based feedstocks will still limit the overall profitability of biofuels. Currently, upwards of half of the production cost of these biofuels needs to be supported by government subsidies (Hill et al., 2006).

The next generation of feedstocks will need to have lower land requirements and lower production cost, yet maintain high production capacity in order to bring biofuels closer to economic viability. Metabolic engineering allows us to bridge the feedstock gap by enabling the utilization of cheaper and more sustainable substrates by introducing catabolic pathways and optimizing metabolic networks for the conversion of feedstock to fuel. Indeed, yield optimization has been a critical aspect of virtually all biochemical engineering processes in recent history. Metabolic engineering of organisms towards this end only serves to continue this tradition, pushing yields beyond what is naturally observed. Furthermore, microbe-based biofuel production also reduces the crop land requirements compared to crop-based methods, decreasing competition with food production.

### 2.1.2 Engineering for Improved Fuels

While new feedstocks are explored, a simultaneous search continues for the next generation of fuel types. Current biofuels have some persistent disadvantages which limit their incorporation into existing infrastructure. Ethanol, while widely produced, has relatively poor fuel characteristics. Ethanol is hygroscopic, capable of absorbing water which can lead to corrosion. The energy content is also low, containing only 70% of the energy per volume of gasoline. Also, as ethanol is produced by fermentation, the resulting beer is dilute, containing roughly 10% ethanol. Subsequent distillation to separate the ethanol is very energy intensive (Pimentel and Patzek, 2005). Biodiesel is a better fuel, but also has some disadvantages. It is not well-suited for use at low temperatures because of a high cloud point, and still often requires large quantities of petroleum-derived methanol as part of its production. It also has only 89% of the energy content of its analog, petrodiesel (Balat and Balat, 2008). Current biofuel characteristics limit their integration into existing infrastructure. Because of this, there is a high transition barrier to adoption of biofuels, and both ethanol and biodiesel are often only blended at low concentrations into conventional fuels.

Development of better fuels that have high energy density and can be integrated into existing pipelines and engines will be needed if biofuels are to be more widely adopted and have a reasonable hope to replace fossil fuels. Through metabolic engineering of production pathways, alternative products can be made that have characteristics closer to their petroleum equivalents, easing the barrier for adoption. These alternatives range from slight modifications to existing metabolites, to new pathways which create naturally unique compounds. These naturally rare products will also require extensive pathway engineering and optimization in order to achieve effective production capacities - one of the central strengths of metabolic engineering.

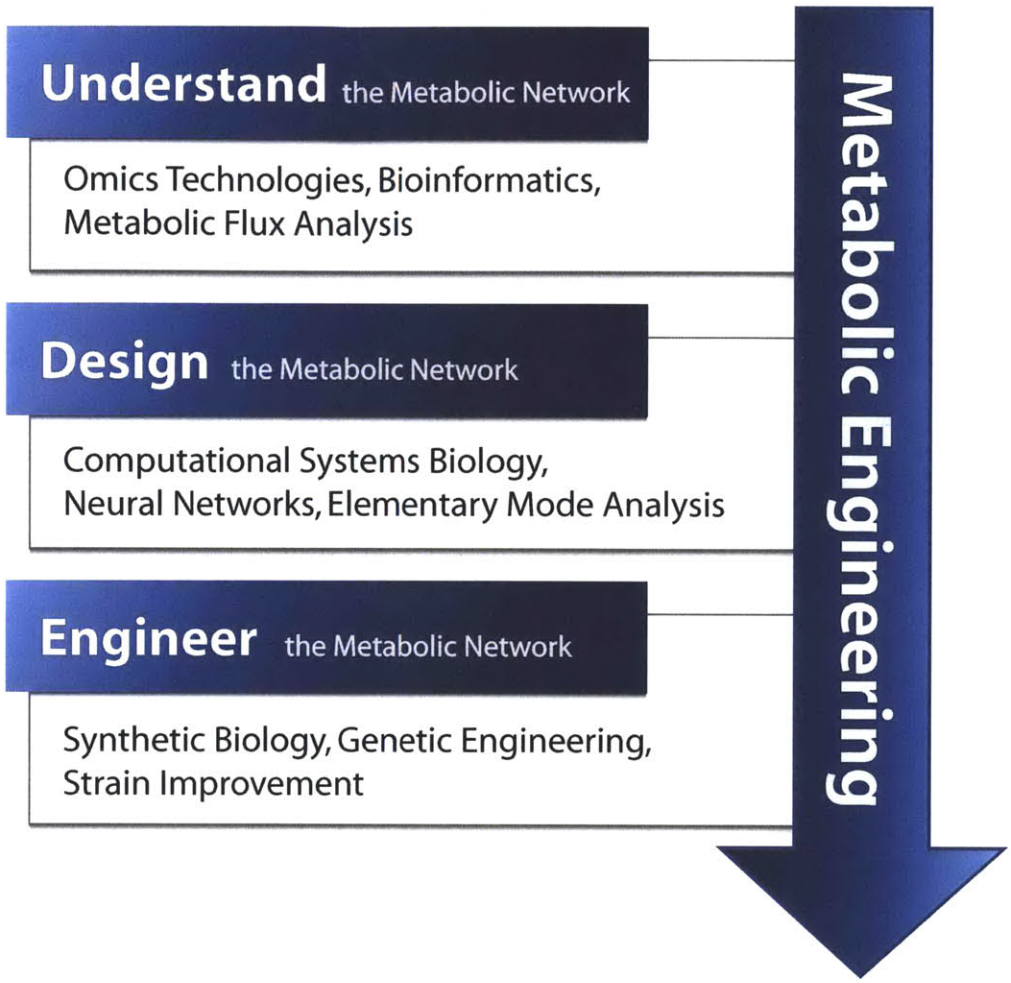


Figure 2.2.1: Metabolic engineering employs techniques and technologies from a range of disciplines: from omics technology to synthetic biology. Strategies revolve around the understanding, design, and engineering of metabolic networks to produce academically and commercially relevant products from biological platforms.

## 2.2 Tools of Metabolic Engineering

To understand how metabolic engineering plays a role in biofuels development and how it takes an interdisciplinary approach to problem solving, it is important to first understand its main strategies and tools. The strategies of metabolic engineering can be compartmentalized into three steps: understanding, designing, and engineering the metabolic network. Each of these steps uses tools and technologies adopted from a range of disciplines. An overview of the strategies of metabolic engineering can be found in Figure 2.2.1.

### 2.2.1 Understanding the Metabolic Network

The first step in metabolic engineering is to understand the complex network of enzymatic reactions which compose a cell's metabolism. In addition to the enzymology of participating enzymes, this requires information on the structure and behavior of the pathways that connect these enzymes. Knowledge of the pathway chemistry and stoichiometry allows us to calculate theoretical yields which are often used as benchmarks for pathway engineering efficacy. Comprehensive systems-level data about these complex networks is acquired through omics technologies and bioinformatics. Omics technologies involve using genomic, transcriptomic, proteomic, or metabolomic data to quantify the system behavior of the cell along various functional axes (e.g. growth, tolerance, productivity) (Joyce and Palsson, 2006). Bioinformatics is the method of extracting biological meaning by identifying significant patterns, motifs, and connections within these large, complex datasets. These techniques enable us to develop a systems-level perspective on cellular activity and an understanding of important contributing networks (Tyo et al., 2007). As an example, metabolic flux analysis derived from metabolomic data allows us to observe the flow of material through cellular metabolic pathways. Like a material balance, these fluxes describe the distribution of material throughout the cell's metabolic network, and can help identify branch points and competing pathways relevant to our desired product. Fluxes also help determine the degree of engagement of various enzymes in the pathway, allowing us to identify rate-limiting steps and control points (Stephanopoulos, 1999). Since any biological manipulation will rarely ever produce only an isolated response, it is important to observe the system-level response of our engineering efforts. Using bioinformatics and omics technologies allows us to understand the interactions, connections and responses between different parts of the system in order to predict and control the metabolic network.

### 2.2.2 Designing the Metabolic Network

Once we have sufficient understanding of the organism and its cellular activities, we are then able to develop and design specific strategies to obtain our desired product. While we can introduce, remove, or otherwise modify pathways, identifying the most effective actions a priori can help save much time and effort. Modern methods to do so are found in the field of computational systems biology. A main goal of computational systems biology is to reconstruct cellular networks *in silico* which can model the behavior of the cell. Starting with a cellular model, one is able to simulate and characterize how possible pathway manipulations will affect the system overall. Evaluation of these changes can help identify the ideal genetic targets that will maximize our objectives. One such method of evaluation is called elementary mode analysis, which uses a systems engineering approach to decompose metabolic networks into uniquely organized pathways that can be used to evaluate cellular phenotypes, metabolic network regulation, network robustness and fragility (Trinh et al., 2009). As an extension, neural networks can also be used to make sense of exceptionally difficult systems and to subsequently predict future behavior (Bosl, 2007).

### 2.2.3 Engineering the Metabolic Network

Once targets and pathways are identified, the next task is to implement these changes *in vivo*. This involves genetic manipulation of the host organism using molecular biology. The term synthetic biology describes the systematic approach to pathway manipulation through standardized biological components for the purpose of increasing their programmability and robustness (Leonard et al., 2008). Under this framework, genetic elements are modularized to simplify the process of genetic engineering. These elements can then be used to introduce new genes, knockout existing genes, or modifying existing DNA sequences. Modules can be built up to produce whole pathways, and can also be rearranged to optimize expression. Numerous non-rational techniques are also available which extend the reach of traditional strain improvement: high throughput screening, directed evolution, gene shuffling, and com-

binatorial engineering. These techniques increase the efficiency of strain improvement by sampling a much larger phenotypic search space, opening up the possibility to select for changes in less-intuitive or distal targets. An interesting example is the use of transcriptional engineering to achieve phenotypic diversity (Alper and Stephanopoulos, 2007). This technique involves mutagenesis of transcription factors to produce global changes in the transcription and expression of genes in the cell, which would not be feasible through isolated point mutations (Alper et al., 2006). Finally optimizing to maximize flux through the production pathway is often the most difficult engineering task. Tuning the expression of genes many times is necessary as intermediates and cofactors need to be balanced within the pathway. Furthermore, troubleshooting the pathway often involves alleviating any number of potential bottlenecks that may impede flux such as competing pathways, lack of enzymatic driving force, cofactor imbalances, insufficient enzyme activity, unbalanced enzyme expression, transport issues, enzyme regulation, and toxicity. Because of its interdisciplinary nature, metabolic engineering will be limited to the available tools and technologies of the fields from which it draws. However as the state of the art progresses and development improves at the interface of these related disciplines, metabolic engineers will have increased power to evaluate, design, and manipulate cells to produce the desired products that will be in demand in the future. These abilities will most certainly be necessary for the development of the next generation of biofuels.

## 2.3 Conclusion

A systems-level interdisciplinary approach is necessary for effective strategies to tackle today's global energy and environmental problems. The tools and strategies of metabolic engineering are well-suited for addressing the persistent challenges facing a successful transition away from petroleum transportation fuels. As such, metabolic engineering will be instrumental in developing the next generation of cost-effective and robust transportation

fuels, which will come from cheaper, more sustainable feedstocks and have better fuel characteristics. It remains to be seen what processes and technologies will successfully establish sustainable alternatives to petroleum transportation fuels. However, because of obstacles in current feedstocks and fuels, it is necessary to continue research in technologies that can overcome existing limitations. Metabolic engineering is uniquely poised to develop and implement the next-generation biofuels using a systems-level approach from multiple disciplines. It enables promising and exciting opportunities for alternative energy, which have the potential for great societal ramifications. Metabolic engineering will also be central in the long road ahead to developing these opportunities into robust, efficient industrial-scale technologies.

# References


Alper, H., Moxley, J., Nevoigt, E., Fink, G. R., and Stephanopoulos, G. (2006) Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production. *Science 314*, 1565–1568.

Alper, H., and Stephanopoulos, G. (2007) Global transcription machinery engineering: A new approach for improving cellular phenotype. *Metab. Eng. 9*, 258–267.

Bailey, J. E. (1991) Toward a science of metabolic engineering. *Science 252*, 1668–1675.

Balat, M., and Balat, H. (2008) A critical review of bio-diesel as a vehicular fuel. *Energy Conversion and Management 49*, 2727–2741.

Bosl, W. (2007) Systems biology by the rules: hybrid intelligent systems for pathway modeling and discovery. *BMC Systems Biology 1*, 13.

Fargione, J., Hill, J., Tilman, D., Polasky, S., and Hawthorne, P. (2008) Land Clearing and the Biofuel Carbon Debt. *Science 319*, 1235–1238.

Hill, J., Nelson, E., Tilman, D., Polasky, S., and Tiffany, D. (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences 103*, 11206.

Joyce, A. R., and Palsson, B. O. (2006) The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell Biol. 7*, 198–210.

Keasling, J. D. (2010) Manufacturing Molecules Through Metabolic Engineering. *Science 330*, 1355–1358.

Leonard, E., Nielsen, D., Solomon, K., and Prather, K. J. (2008) Engineering microbes with synthetic biology frameworks. *Trends Biotechnol. 26*, 674–681.

Pimentel, D., and Patzek, T. W. (2005) Ethanol Production Using Corn, Switchgrass, and Wood; Biodiesel Production Using Soybean and Sunflower. *Natural Resources Research 14*, 65–76.

Stephanopoulos, G. (2007) Challenges in engineering microbes for biofuels production. *Science 315*, 801.

Stephanopoulos, G. (1999) Metabolic Fluxes and Metabolic Engineering. *Metab. Eng. 1*, 1–11.

Trinh, C., Wlaschin, A., and Srienc, F. (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl. Microbiol. Biotechnol. 81*, 813–826.

Tyo, K. E., Alper, H. S., and Stephanopoulos, G. N. (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol. 25*, 132–137.

# Chapter 3

# *Yarrowia lipolytica*: Platform for Yeast Oil Production

## 3.1 General Remarks

Metabolic engineering for microbial biodiesel production requires a platform for yeast oil production. The use of oleaginous microorganisms provides advantages in that these organisms naturally accumulate and process lipids, and potentially have strong and robust lipid biosynthesis pathways. The oleaginous yeast, *Yarrowia lipolytica* (formerly *Candida lipolytica*) is a dimorphic filamentous fungus that is both capable of consuming and accumulating large amounts of lipids. It is one of the most extensively studied "non-conventional" yeasts, with applications in both industrial and basic research. It is an obligate aerobe and is considered non-pathogenic, carrying the designation as generally recognized as safe (GRAS) by the Food and Drug Administration (FDA). It is often isolated from foods, like cheese, but can also be found in the ocean, soil, and other natural environments rich in oils. The genome of *Y. lipolytica* was first sequenced in 2004, spanning six chromosomes totaling 20.5 Mb of DNA. Phylogenetic analysis reveals that *Y. lipolytica* is distantly related to *Saccharomyces cerevisiae*, although taxonomically they are both classified in the hemiascomycetous clade.

*Y. lipolytica* has been used extensively as a model organism for a wide range of biological studies. These studies range from fundamental investigation into characteristic aspects of its physiology to studies to harness *Y. lipolytica* for possible industrial applications. *Y. lipolytica*, as an obligate aerobe, requires oxygen for growth. The dependence and evolution of a strong and robust respiratory pathway has allowed for a number of studies on the energetics of growth in *Y. lipolytica* and mitochondrial characterization. Additionally, the presence of an alternative pathway for oxidative phosphorylation has highlighted its unique nature in the fungal kingdom. As a dimorphic filamentous fungus, *Y. lipolytica* can transition between two distinct morphologies depending on environmental conditions: elongated tube-like mycelia and round spherically-shaped yeast. This phenomenon has been linked to virulence in a wide range of pathogenic fungi. Research on the complex mechanism and triggers associated with dimorphism has uncovered a variety of influences, including pH, carbon and nitrogen sources. Typically cultures will consist of a mixed population of both hyphal and yeast cells.

One gene associated with transition to mycelium, the deletion of which abolishes mycelium formation, is the HOY1 gene.

*Y. lipolytica* has also been used as a model organism for lipid accumulation. As an oleaginous yeast, *Y. lipolytica* can naturally accumulate up to 40% lipids in the form of intracellular lipid bodies, predominantly composed of triacylglycerides (TAG). TAG accumulation can occur from either exogenous oil uptake or *de novo* biosynthesis from sugar substrates. Lipid accumulation is triggered by nutrient limitation in the presence of excess carbon. In laboratory settings, the lipid accumulation is most commonly induced through limitation of nitrogen. Upon depletion of nitrogen, both transcriptome and proteome-level modifications serve to modulate a variety of pathways and divert the flux of carbon towards lipid synthesis and accumulation. The molar ratio between carbon and nitrogen (C/N) is an important factor in the determination of oil-production potential of the culture, as the ratio helps determine the amount of excess carbon in the environment once nitrogen has been depleted.

*Y. lipolytica* naturally secretes proteins in response to extracellular conditions. In the presence of extracellular peptides, *Y. lipolytica* will secrete a variety of proteases depending on the pH of the environment. These acid and alkaline proteases are then used to promote utilization of the extracellular peptides. In the presence of lipids, oils and hydrophobic carbon substrates, *Y. lipolytica* will secrete an array of lipases to digest and promote uptake of these substrates for intracellular utilization and catabolism. In addition to mechanistic studies of the protein secretion pathway, both protein products have commercial value and have been seen in industrial settings and applications.

One of the most prominent metabolic products of *Y. lipolytica* is secreted citric acid. As an industrially and commercially relevant organic acid, it has been studied as a platform for citric acid production. Excess citrate is often produced in conditions similar to those used in lipid production: excess carbon with limiting nitrogen. Additionally, *Y. lipolytica* has been studied for its production of citric acid on non-conventional substrates: raw glycerol,

glucose hydrolysate, olive-mill wastewater. These applications exploit the natural ability of *Y. lipolytica* to assimilate and tolerate a wide-range of carbon substrates and environmental conditions. From these studies, *Y. lipolytica* has proven itself to be a robust organism with industrially-relevant applications. Additionally, basic research into the physiology, genomic sequence, and genetic tools for the manipulation of *Y. lipolytica* have vastly accelerated its usefulness and potential.

## 3.2 Metabolism

While *Y. lipolytica* and *S. cerevisiae* both are hemiascomycetous yeast, they are possibly the most evolutionarily divergent inside the group (Gaillardin et al., 2000; Bon et al., 2003; Blank et al., 2005). Thus it is not surprising that the metabolism of *Y. lipolytica* often differs significantly from the model organism *S. cerevisiae*. These differences result in distinguishing characteristics identifying *Y. lipolytica* as oleaginous, Crabtree-negative and an obligate aerobe. This section will characterize the metabolism of *Y. lipolytica*, particularly by highlighting some observed similarities and differences with the metabolism of the model organism *S. cerevisiae*.

### 3.2.1 Carbon Metabolism

#### Hexose

*Y. lipolytica* can readily catabolize a variety of hexoses, in particular glucose, fructose and mannose. These sugars enter into the cell via specific transmembrane proteins called either transporters or permeases. Glucose uptake is handled by two systems that are regulated by glucose levels: high-affinity transport and low-affinity transport. High-affinity transporters have low $K_m$, meaning they are readily active at low substrate concentration - generally speaking these transporters more commonly require ATP. Low-affinity transporters have high $K_m$, which requires higher concentrations to activate but typically uptake at higher

rates, utilizing concentration gradients for transport. Furthermore, the mechanism for glucose transport can differ between organisms. *S. cerevisiae* contains almost 20 HXT hexose transporters, all of which utilize facilitated diffusion as the transport mechanism (Maier et al., 2002). This involves non-energetic conformational changes in transmembrane proteins to shuttle the solute across the membrane; no ATP is consumed. Many of the transporters in *Y. lipolytica* show similarity to GXS1 glucose transport protein, which belongs to the fungal monosaccharide-$H^+$ symporter family (Leandro et al., 2006). These systems transport protons along with the substrate, and ATP is subsequently required to remove the protons and maintain pH balance. This suggests that *Y. lipolytica* utilizes an active transport of glucose and thus consumes ATP for its hexose uptake. *Y. lipolytica* transport also does not show any glucose repression or concentration-dependence for carrier activity, resulting in similar uptake rates at both high and low concentrations (Does and Bisson, 1989). Glucose uptake in *Y. lipolytica* was found to be about one-fourth that of *S. cerevisiae* (Does and Bisson, 1989), and glycolytic rates were similarly low (Blank et al., 2005).

### Xylose

Despite the presence of open reading frames similar to xylose reductase and xylitol dehydrogenase found its genome, *Y. lipolytica* has been classified as an organism that cannot grow on xylose (Pan et al., 2009). The biochemical characteristics of xylose transport in general are still not well understood (Leandro et al., 2006, 2009). Xylose consumption would allow *Y. lipolytica* to be able to grow off hemicellulose substrates, making it a potential means of producing cellulosic biodiesel. However, xylose consumption and lipid production are potentially at odds. Xylose consumption requires high PPP flux, which *Y. lipolytica* does have (Blank et al., 2005), but lipid accumulation begins as the PPP flux is deactivated (nitrogen-limiting conditions limit need for ribose) (Zhang et al., 2007). Also the regulatory networks associated with lipid accumulation with respect to xylose consumption are not well investigated (the cell may be doing very unconventional things because of this configuration).

A study of *Aspergillus niger*, a cellulolytic mold, found that up to 16.6% lipid accumulation occurred growing off various substrates including xylose and bagasse (Singh, 1992). Other oleaginous yeasts have been shown to grow off xylose, assimilating xylose and accumulating triglycerides under nitrogen-limiting conditions, yielding approximately 0.15 g lipid/g xylose (Fall et al., 1984; Pan et al., 2009). The theoretical maximum yield for lipid production from xylose is 0.34 g lipid/g xylose (Ratledge, 1988). Because xylose typically elicits a respiratory response (Jin et al., 2004), and also because xylose uptake enters through the pentose phosphate pathway rather than the glycolytic pathway, many of the lipid synthesis pathways may not be triggered in the presence of this sugar, resulting in decreased experimental lipid accumulation and yield.

### Hydrophobic Substrates

Perhaps one of the most unique characteristics of *Y. lipolytica* is its ability to utilize a wide range of hydrophobic substrates, from lipids and fatty acids to n-alkanes and 1-alkenes (Barth and Gaillardin, 1997). While fatty acids can readily be assimilated into the organism, lipids and hydrocarbons require extracellular processing in the form of lipases and emulsifiers, which break down and facilitate adsorption onto the cell surface and subsequent incorporation and utilization.

Fatty acids are introduced into the cell and transported to the peroxisome where they are converted into the corresponding acyl-coA molecules and subsequently degraded by peroxisomal $\beta$-oxidation into acetyl-coA and propionyl-coA (in the case of odd-chain molecules). Since fatty acid synthesis is repressed in the presence of exogenous fatty acids, the acetyl-coA is converted into TCA cycle intermediates via the glyoxylate shunt. Further anaplerosis is possibly via gluconeogenesis.

On the other hand, because alkanes and alkenes typically lack any reactive functional groups, the catabolism of these hydrocarbons first requires the use of cytochrome P-450 enzymes found on the endoplasmic reticulum to hydroxylate the molecule. The subsequent

1-alkanol is then oxidized by a membrane-bound fatty alcohol oxidase to form fatty acids that can be catabolized normally (Coelho et al., 2010).

### Alcohols

While *Y. lipolytica* typically will not generate ethanol as an obligate aerobe, it can consume it as a sole carbon source up to concentrations of 3%. A number of $NAD^+$ and $NADP^+$-dependent alcohol dehydrogenases have been observed (Barth and Gaillardin, 1997). Higher concentrations were found to be toxic to the cell. As the ICL promoter is induced by ethanol and is also necessary for growth on ethanol (Juretzek et al., 2001; Flores and Gancedo, 2005), it is likely that ethanol is catabolized through the glyoxylate shunt, similar to utilization of fatty acids and acetate.

Glycerol may also be used as a carbon substrate, and can be assimilated via two pathways: dihydroxyacetone and glycerol-3-phosphate pathways. In *Y. lipolytica*, it is thought to predominantly go through the latter route, where glycerol is converted by glycerol kinase to form glycerol-3-phosphate (Beopoulos et al., 2008; Morgunov and Kamzolova, 2011). The glycerol-3-phosphate can then be converted into DHAP for use in central metabolism, or be used as a backbone for TAG assembly predominantly through the Kennedy pathway (Beopoulos et al., 2008).

### Acetate

*Y. lipolytica* has been found to grow efficiently on acetate as sole carbon source. In literature, concentrations up to 0.4% of sodium acetate are well tolerated, while higher concentrations reduce the growth rate and concentrations above 1.0% inhibit growth. However, very little is known about the uptake of acetate by *Y. lipolytica*. One protein, Gpr1, has been identified to aid adaptation to acetic acid (Augstein et al., 2003). It is thought that protonated acetic acid freely diffuses through the plasma membrane, and Gpr1 regulates both weak acid efflux pumps and/or plasma membrane $H^+$-ATPase to eject the proton once the acetic acid has

dissociated and the acetate has been consumed. The deletion of Gpr1 thus inhibits the intracellular pH control and results in acetic acid sensitivity.

### 3.2.2 Respiration

The respiratory chain of *Y. lipolytica* is more similar in function to higher eukaryotes, like plants and mammals, than fermentative yeast such as *S. cerevisiae*. As an obligate aerobe, cellular growth is coupled to the presence of oxygen and function of cellular respiration. Numerous studies have been performed modeling and optimizing oxygen transfer and aeration in bioreactor systems for *Y. lipolytica*. Additionally, as a Crabtree-negative organism, *Y. lipolytica* does not divert excess carbon flux through fermentative pathways in the presence of excess glycolytic flux. As a result a vast majority of catabolic carbon flux passes into the mitochondria and into the TCA cycle (Blank et al., 2005).

*Y. lipolytica* uniquely contains two "alternative" enzymes that augment the respiratory chain: NDH2 and AOX. NDH2 is a NADH:ubiquinone oxidoreductase that carries out the same redox reaction as the normal mitochondrial complex I, but is not coupled to proton transport. Thus, the oxidation of NADH via this alternative pathway would not contribute to the proton-motive force in the mitochondria. These alternative enzymes thus seem to form a futile oxidation pathway and are responsible for a cyanide-resistant respiration phenotype (Kerscher et al., 2002). It has been suggested that the alternative oxidase is used to mitigate radical oxygen species production from mitochondrial complex I activity, particularly in stationary phase (Guerrero-Castillo et al., 2012).

### 3.2.3 Pentose Phosphate Pathway

During glycolysis, a portion of flux is diverted towards the PPP primarily for three purposes: (1) generate reducing equivalents in the form of NADPH, used for fatty acid synthesis and biomass generation (2) generate ribose-5-phosphate, used in nucleotide synthesis, and (3) generate erythrose-4-phosphate, used in the synthesis of aromatic amino acids. The relative

amount of flux entering the PPP can often be an indication of the metabolic prioritizations of the organisms, as the PPP provides important precursors for biomass generation. The relative flux towards the PPP has been positively correlated with biomass yield (Blank et al., 2005). While the rate of glucose uptake and flux through glycolysis is roughly four-fold lower than *S. cerevisiae*, *Y. lipolytica* interestingly exhibits a much higher relative flux through the pentose phosphate pathway (PPP). Comparative metabolic flux analysis showed that 10% of glycolytic flux is diverted towards the PPP in *S. cerevisiae*, while 40% glycolytic flux is diverted in *Y. lipolytica* (Blank et al., 2005).

### 3.2.4 Catabolite Repression

Catabolite repression is the regulatory mechanism for an organism to prioritize the utilization of a specific preferred substrate over other alternative substrates by repressing alternative pathways in the presence of preferred substrates. In the case of glucose repression, the presence of glucose acts to downregulate the expression of transporters and pathways that are used to consume other substrates. The result is the sole consumption of glucose before any other substrate. In situations of cofermentation of multiple substrates, a diauxic shift is observed as the organism first completely consumes a preferred substrate, then transitions to the uptake and catabolism of the other substrate. A study of the substrate cofermentation by *Y. lipolytica* identified the following preferences (in order of most preferred): hexadecane, glucose, and oleate and glycerol (Morgunov and Kamzolova, 2011). Furthermore, repression may not be absolute, allowing some substrate preferences overlap: oleate uptake initiates even in the presence of low concentrations of glucose (<2.5 g/L); glycerol and oleate are actually consumed simultaneously. Interestingly, *Y. lipolytica* can be found to exhibit incomplete glucose repression of certain pathways. Pyruvate carboxylase (PYC) is the main anaplerotic pathway in high-glucose conditions since the alternative glyoxylate shunt is normally repressed in glucose-rich environments. Under glucose conditions, ΔPYC mutants normally can't grow, since the glyoxylate shunt is repressed. However, *Y. lipolytica* displays

an unusual growth phenotype, evidently because the glyoxylate shunt is still active.ΔPYC ΔICL (isocitrate lyase, first step in the glyoxylate shunt) double mutant cannot grow (Flores and Gancedo, 2005). Interestingly, one of the first examples of metabolic flux analysis ever published was on *Y. lipolytica* (Aiba and Matsuoka, 1979). In the study, both pyruvate carboxylase and glyoxylate fluxes were necessary to calculate a solution, indicating non-zero fluxes in both. The incomplete glucose repression found with the glyoxylate shunt exemplifies the differences in regulatory control and feedback in *Y. lipolytica* compared to the more commonly studied *S. cerevisiae.* Indeed lipid synthesis, which is commonly understood as a tightly regulated pathway in most organisms, may have a very different regulatory profile in *Y. lipolytica* contributing to its oleaginicity.

## 3.3 Lipid Biosynthesis

As an oleaginous yeast, the lipid metabolism of *Y. lipolytica* includes a few unique features that encourage robust lipid synthesis. The complete metabolic pathway for production of lipids from glucose is depicted in Figure 3.3.1. This section surveys the pathways, metabolites, and enzymes that are component to lipid biosynthesis.

### 3.3.1 Glycolysis

Glucose and most carbohydrate substrates enter central metabolism through glycolysis. This process encompasses the catabolism of sugar to pyruvate. *De novo* synthesis of glycerol-3-phosphate, which acts as the backbone for TAG synthesis, can be generated from the reduction of dihydroxyacetone phosphate (DHAP) via glycerol-3-phoasphate dehydrogenase with an NADH cofactor. Pyruvate, on the other hand, enters the mitochondria to enter the tricarboxylic acid (TCA) cycle. Since *Y. lipolytica* is a Crabtree-negative respiratory organism (i.e. it will not ferment sugar), a large amount of relative carbon flux enters the TCA cycle - approximately 65% (Blank et al., 2005). Consequently, pyruvate then under-

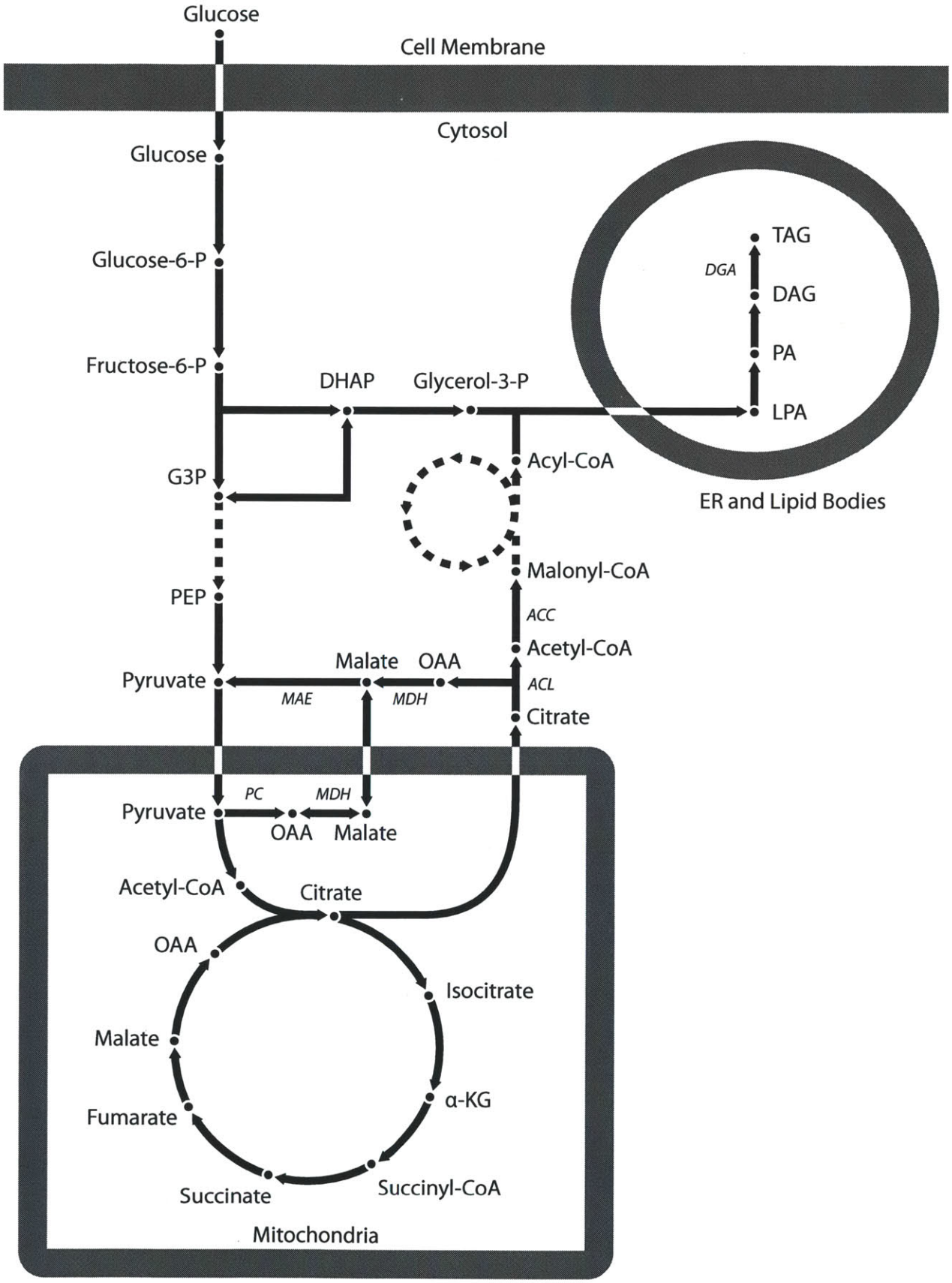


Figure 3.3.1: Metabolic pathway for lipid production from glucose

goes decarboxylation to produce mitochondrial acetyl-CoA via the pyruvate dehydrogenase complex (PDC).

### 3.3.2 Citrate Shuttle

Mitochondrial acetyl-CoA then enters the TCA cycle proper through the condensation with oxaloacetate to form citrate. At this point, mitochondrial citrate can be used in two pathways. The citrate can be converted to isocitrate to participate in the TCA cycle, where it can be oxidized to generate the reducing equivalent NADH for use in oxidative phosphorylation to generate ATP. This process of respiration is the primary energy generating mechanism in these organisms. Alternatively, the citrate can exit the mitochondria and be cleaved into oxaloacetate and cytosolic acetyl-CoA.

ATP citrate lyase (ACL) is the cytosolic enzyme that cleaves the exported citrate into acetyl-coA and oxaloacetate in the cytosol. This acetyl-coA can then be utilized as a precursor for fatty acid synthesis. This enzyme has been identified as a key distinction between oleaginous and non-oleaginous yeasts, being differentially present in the former and not the latter (Boulton and Ratledge, 1981). *S. cerevisiae* does not have a citrate lyase; it utilizes an acetate pathway to generate acetyl-CoA in the cytosol for fatty acid synthesis. *Y. lipolytica* does have ATP:citrate lyase, allowing for the utilization of citrate as a carrier for acetyl-groups from the mitochondria to the cytosol. However, most cross-species analysis shows that while ACL is necessary for lipid accumulation, the level of expression is not strongly influential of final lipid content, nor is it correlated with degree of oleaginicity. One study cloned RTACLY into tobacco and achieved only a 16% increase in lipid accumulation (Ratledge and Wynn, 2002).

The remaining oxaloacetate can subsequently be converted into malate or pyruvate to reenter the mitochondria. The net result of this action is the transport of mitochondrial acetyl-CoA to cytosolic acetyl-CoA, commonly known as the citrate shuttle. This shuttle is the main source of cytosolic acetyl-CoA in oleaginous yeast as well as higher eukaryotes.

### 3.3.3 Fatty Acid Synthesis

Cytosolic acetyl-CoA is the primary precursor for fatty acid synthesis. Fatty acid synthesis is the process by which carbon from acetyl-CoA is cyclically condensed to form large fatty acid molecules. Acetyl-CoA is carboxylated to form malonyl-CoA via the activity of acetyl-CoA carboxylase (ACC). Malonyl-CoA is then used as the monomer in a decarboxylation reaction for the elongation cycle performed in the fatty acid synthase (FAS) complex. The elongating molecule in the FAS complex is not a CoA molecule; instead it is derivatized with an acyl-carrier protein (ACP), which facilitates movement within the complex. However, once the chain has been finally extended to the preferred length, it is then released and esterified with free CoA to reform an acyl-CoA molecule. In eukaryotes, the preferred length and CoA derivatization of the fatty acid synthase complex results in cytosolic generation of palmitoyl-CoA (C16).

### 3.3.4 NADPH Generation

During fatty acid synthesis, besides the consumption of acetyl-CoA, NADPH is also necessary to provide the reducing equivalents necessary to reduce acyl-CoA molecule. There are two possible sources of cytosolic NADPH: the pentose phosphate pathway, and the transhydrogenase cycle.

In many oleaginous yeast, it has been suggested that the transhydrogenase cycle, via the activity of malic enzyme, primarily produces the NADPH reducing equivalents for fatty acid synthesis (Wynn et al., 1999). This was suggested after studies of *Aspergillus nidulans* and *Mucor circinelloides* found that the disruption of malic enzyme activity abolished lipid accumulation (Wynn et al., 1997; Wynn and Ratledge, 1997). Additionally, overexpression of malic enzyme resulted in a 2.5-fold increase in lipid accumulation in *Mucor circinelloides* (Zhang et al., 2007). Furthermore, it has been suggested that in oleaginous microorganisms, since nitrogen is depleted in lipid accumulation phases, the pentose phosphate pathway is naturally down-regulated, since protein and nucleic acid production can no longer proceed

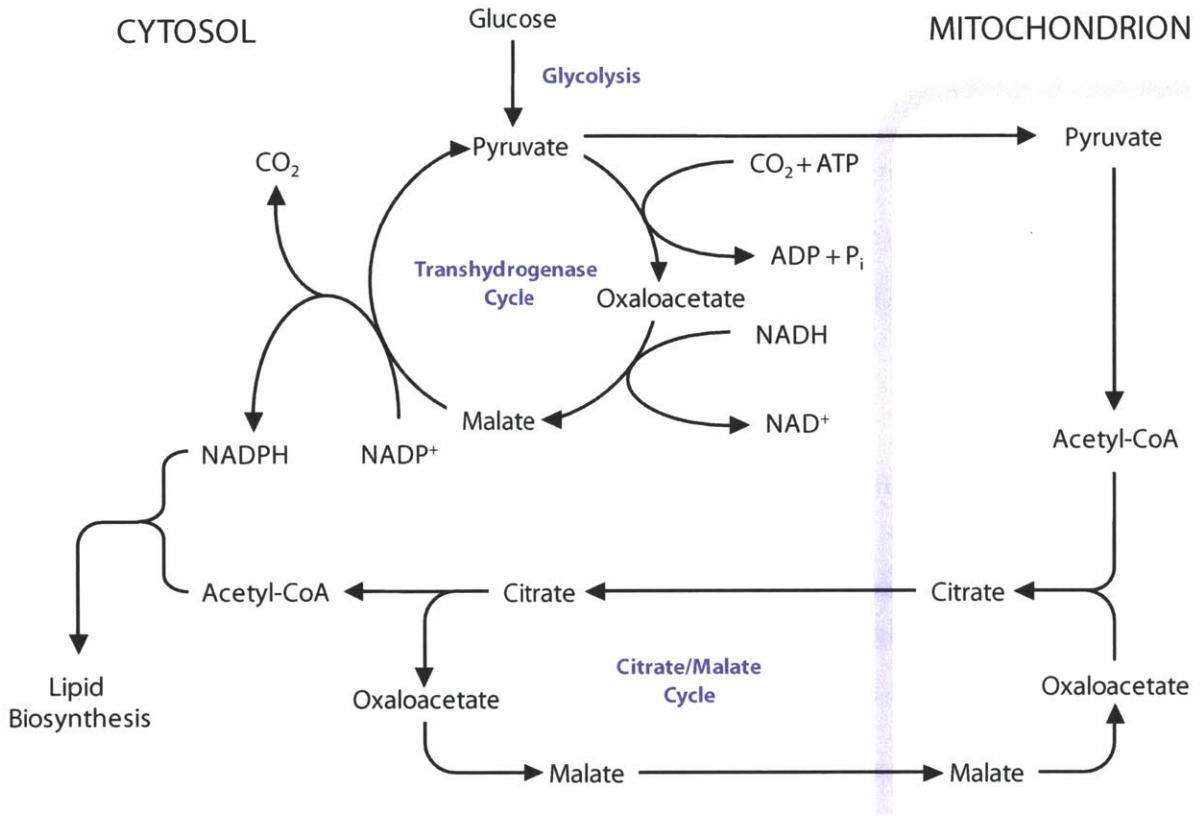


Figure 3.3.2: Transhydrogenase cycle and Citrate shuttle, adapted from Ratledge (2004).

and therefore there is no need for pentose. Thus malic enzyme is particularly useful for providing the necessary NADPH during lipid accumulation.

As malic enzyme activity has been observed in a wide number of oleaginous yeasts, it has also been proposed that malic enzyme participates in an integrated metabolon complex where it is physically scaffolded with ATP citrate lyase and the fatty acid synthase complex, facilitating channeling of acetyl-CoA and NADPH into fatty acid production (Ratledge, 2004). However, currently there is no direct experimental evidence of the integrated metabolon complex existing.

For *Y. lipolytica*, the situation is less concrete, as more lines of evidence point towards NADPH generation coming from the pentose phosphate pathway rather than through malic enzyme. From metabolic flux experiments, the relatively high flux of carbon through the pentose phosphate pathway would seem to be more than sufficient to provide NADPH for lipid accumulation (Blank et al., 2005). It was also found that there is also very little flux from malate to pyruvate via malic enzyme in both linear phase and stationary phases (Walther, 2010). Furthermore, unpublished data from Beopoulos et al. found that malic enzyme overexpression did not affect lipid accumulation (2011).

### 3.3.5 Elongation and Desaturation

Once fatty acid synthesis has produced palmitoyl-CoA in the cytosol, the organism can further modify the fatty acid depending on cellular needs and/or preferences. Cytosolic palmitoyl-CoA will migrate to the endoplasmic reticulum (ER), where it partitions into the ER membrane. On the membrane surface, additional reactions can occur to either elongate or desaturate the acyl-CoA. These two processes are respectively mediated by fatty acid elongase and desaturase enzymes. In *Y. lipolytica*, there are strong preferences for fatty acid production of palmitate (C16:0) and oleate (C18:1).

### 3.3.6 TAG Assembly

Long chain fatty acyl-CoA molecules can be used in a variety of pathways: they can be incorporated into the cell membrane, be used as signaling molecules, or be stored for later energy generation. Storage of fatty acids is especially strong in oleaginous yeasts and during lipid accumulation, as in these situations, fatty acids may accumulate in great number because of metabolism or environmental conditions. In *Y. lipolytica*, fatty acids are assembled into TAG for lipid storage and accumulation via the Kennedy pathway.

While all TAG synthesis pathways ultimately combine three acyl-CoA monomers to a glycerol backbone, the Kennedy pathway utilizes glycerol-3-phosphate as the backbone and condenses the acyl-CoA monomers to first form lysophosphatidic acid (LPA) and then phosphatidic acid (PA). The phosphate is then removed to form diacylglycerol (DAG), and then a final acyl-CoA is added to form TAG. LPA, PA, DAG can also be used in phospholipid synthesis, so in some situations, there is competition for these molecules between storage and membrane synthesis.

### 3.3.7 Lipid Body Formation

TAG assembly ultimately results in the storage of TAG in lipid bodies. In *Y. lipolytica*, analysis of lipid bodies found that roughly 80% of the organelle mass is TAG, about 10% is ergosteryl ester, 5% protein, and the remainder as ergosterol, steryl esters and phospholipids (Athenstaedt et al., 2006). The formation of lipid bodies involves the budding of concentrated TAG from the surface of the ER. The ER membrane surface is used as the phospholipid monolayer that encapsulates the lipid vesicle, and proteins from the surface of the ER are also found on the lipid body membrane (Athenstaedt and Daum, 2006). Since the lipid body retains many of the ER proteins that are used for fatty acid elongation, desaturation, and TAG assembly, the lipid body itself is still capable of both lipid storage and mobilization, independent of the ER. Indeed, lipid storage is best thought of as a dynamic balance between storage and degradation that can rapidly respond to the energy needs of

Figure 3.3.3: Microscope image of lipid bodies in *Yarrowia lipolytica*. Lipid body formation can occur in both hyphael and yeast forms.

the cell. The balancing action of lipid body membrane-bound diacyglycerol acyltransferases and triacylglycerol lipases helps establish this equilibrium (Athenstaedt et al., 2006). A microscope image of lipid bodies in *Y. lipolytica* is shown in Figure 3.3.3.

### 3.3.8 Regulation

As an oleaginous yeast, the cell metabolism is well geared towards the activation and upregulation of lipogenic machinery. A number of key regulators in lipid synthesis have been found in eukaryotes, often involving energy regulation and signalling pathways, as well as transcriptional and allosteric control of rate-limiting steps. The enzyme AMP Kinase (AMPK) monitors the energy state of the cell and responds with the phosphorylation of a broad array of enzymes to regulate anabolic and catabolic pathways for the utilization and storage of energy. Since lipid synthesis is intimately associated with storage and utilization of energy, it is strongly regulated by this enzyme. In states of low energy (low ATP/high AMP concentra-

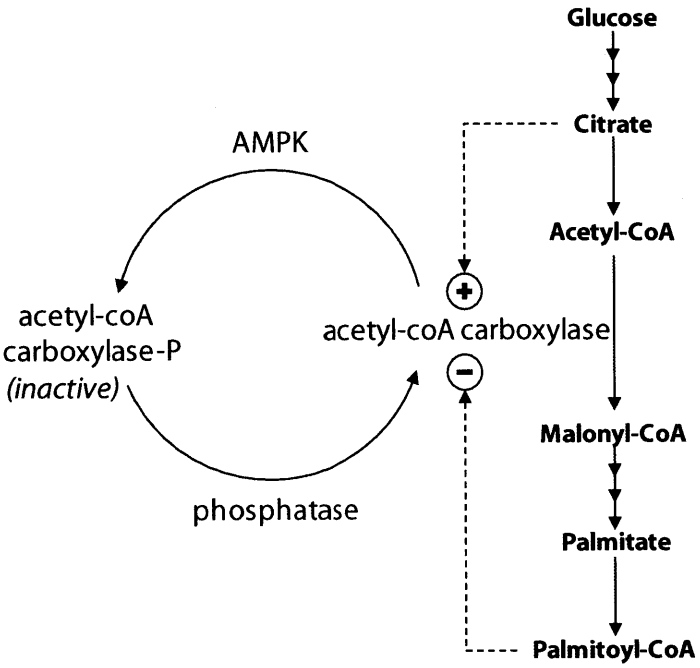


Figure 3.3.4: Feedback regulation of acetyl-coA carboxylase enzyme. Citrate, a precursor of fatty acid synthesis, allosterically activates ACC, while palmitoyl-coA, the product of fatty acid synthesis, inhibits ACC activity. ACC can also be activated and deactivated through phosphorylation by the global energy regulator, AMPK.

tion) or cell stress, pathways to regenerate ATP like fatty acid $\beta$-oxidation and glycolysis are upregulated while pathways consuming ATP are inhibited (e.g. triglyceride and cholesterol synthesis). In states of high energy, the opposite occurs, with AMPK promoting lipogenesis for the storage of excess energy (Hardie and Pan, 2002). This enzyme is well-conserved from yeast to humans and is often described as a global energy regulator and even as a cellular fuel gauge (Hardie and Carling, 1997). In yeast, this enzyme is typically expressed by the SNF gene family. In *S. cerevisiae*, deletion of snf2 resulted in an oleaginous phenotype, accumulating greater than 20% lipid content (Kamisaka et al., 2007).

In addition to global regulation of lipid synthesis, control of the activity of ACC also is used to regulate lipid metabolism. As the first committed step of fatty acid synthesis, downregulation or deactivation of this enzyme results in commensurate deactivation of the entire pathway. In most organisms, ACC is directly phosphorylated by AMPK; the phosphorylation of ACC inhibits enzymatic activity. Additionally, ACC is allosterically regulated by

a number of metabolites for both postive and negative feedback control: citrate activates the enzyme, acyl-coA (typically palmityl-coA) deactivates the enzyme (Ohlrogge and Jaworski, 1997). A number of allosteric inhibitors controlling flux through ACC is depicted in Figure 3.3.4.

## 3.4 Conclusion

The oleaginous yeast *Y. lipolytica* represents a strong candidate as a platform for understanding and increasing lipid production for a microbial bioprocess. It has been widely studied for a variety of biological and industrial aspects and has shown itself to be a robust organism capable of high flux towards lipid synthesis. The study of *Y. lipolytica* and oleaginous yeast in general has shed a great deal of light on the physiology, metabolism and pathways associated with lipid biosynthesis. With our current understanding of its metabolism, we are well-poised to utilize metabolic engineering towards the overproduction of yeast oils.

# References

Aiba, S., and Matsuoka, M. (1979) Identification of metabolic model: Citrate production from glucose by *Candida lipolytica*. *Biotechnol. Bioeng. 21*, 1373–1386.

Athenstaedt, K., and Daum, G. (2006) The life cycle of neutral lipids: synthesis, storage and degradation. *Cell Mol Life Sci 63*, 1355–1369.

Athenstaedt, K., Jolivet, P., Boulard, C., Zivy, M., Negroni, L., Nicaud, J. M., and Chardot, T. (2006) Lipid particle composition of the yeast *Yarrowia lipolytica* depends on the carbon source. *Proteomics 6*, 1450–1459.

Augstein, A., Barth, K., Gentsch, M., Kohlwein, S. D., and Barth, G. (2003) Characterization, localization and functional analysis of Gpr1p, a protein affecting sensitivity to acetic acid in the yeast Yarrowia lipolytica. *Microbiology 149*, 589–600.

Barth, G., and Gaillardin, C. (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. *FEMS Microbiol. Rev. 19*, 219–237.

Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M. T., Hapala, I., Papanikolaou, S., Chardot, T., and Nicaud, J. M. (2008) Control of lipid accumulation in the yeast *Yarrowia lipolytica*. *Appl. Environ. Microbiol. 74*, 7779.

Beopoulos, A., Nicaud, J.-M., and Gaillardin, C. (2011) An overview of lipid metabolism in yeasts and its impact on biotechnological processes. *Appl. Microbiol. Biotechnol. 90*, 1193–1206.

Blank, L. M., Lehmbeck, F., and Sauer, U. (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res. 5*, 545–558.

Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuvéglise, C., Munsterkotter, M., Guldener, U., Mewes, H. W., Van Helden, J., Dujon, B., and Gaillardin, C. (2003) Molecu-

lar evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res. 31*, 1121–1135.

Boulton, C. A., and Ratledge, C. (1981) Correlation of Lipid Accumulation in Yeasts with Possession of ATP: Citrate Lyase. *Journal of General Microbiology 127*, 169–176.

Coelho, M. A. Z., Amaral, P. F. F., and Belo, I. In *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*; Mendez-Vilas, A., Ed.; Formatex Research Center, 2010; Vol. 2; Chapter Yarrowia lipolytica: an industrial workhorse, pp 930–944.

Does, A. L., and Bisson, L. F. (1989) Comparison of glucose uptake kinetics in different yeasts. *J. Bacteriol. 171*, 1303–1308.

Fall, R., Phelps, P., and Spindler, D. (1984) Bioconversion of xylan to triglycerides by oil-rich yeasts. *Appl. Environ. Microbiol. 47*, 1130–1134.

Flores, C.-L., and Gancedo, C. (2005) *Yarrowia lipolytica* Mutants Devoid of Pyruvate Carboxylase Activity Show an Unusual Growth Phenotype. *Eukaryotic Cell 4*, 356–364.

Gaillardin, C. et al. (2000) Genomic Exploration of the Hemiascomycetous Yeasts: 21. Comparative functional classification of genes. *FEBS Lett. 487*, 134–149.

Guerrero-Castillo, S., Cabrera-Orefice, A., Vázquez-Acevedo, M., González-Halphen, D., and Uribe-Carvajal, S. (2012) During the stationary growth phase, *Yarrowia lipolytica* prevents the overproduction of reactive oxygen species by activating an uncoupled mitochondrial respiratory pathway. *Biochimica et Biophysica Acta (BBA) - Bioenergetics 1817*, 353–362.

Hardie, D. G., and Carling, D. (1997) The AMP-activated protein kinase–fuel gauge of the mammalian cell? *Eur J Biochem 246*, 259–273.

Hardie, D. G., and Pan, D. A. (2002) Regulation of fatty acid synthesis and oxidation by the AMP-activated protein kinase. *Biochem Soc Trans 30*, 1064–1070.

Jin, Y.-S., Laplaza, J. M., and Jeffries, T. W. (2004) *Saccharomyces cerevisiae* Engineered for Xylose Metabolism Exhibits a Respiratory Response. *Appl. Environ. Microbiol. 70*, 6816–6825.

Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G., and Nicaud, J. (2001) Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica. Yeast 18*, 97–113.

Kamisaka, Y., Tomita, N., Kimura, K., Kainou, K., and Uemura, H. (2007) DGA1 (diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the δsnf2 disruptant of *Saccharomyces cerevisiae. Biochemal Journal 408*, 61–68.

Kerscher, S., Dröse, S., Zwicker, K., Zickermann, V., and Brandt, U. (2002) Yarrowia lipolytica, a yeast genetic system to study mitochondrial complex I. *Biochim Biophys Acta 1555*, 83–91.

Leandro, M. J., Fonseca, C., and Gonçalves, P. (2009) Hexose and pentose transport in ascomycetous yeasts: an overview. *FEMS Yeast Res. 9*, 511–525.

Leandro, M. J., Gonçalves, P., and Spencer-Martins, I. (2006) Two glucose/xylose transporter genes from the yeast Candida intermedia: first molecular characterization of a yeast xylose–H+ symporter. *Biochem. J. 395*, 543–549.

Maier, A., Völker, B., Boles, E., and Fuhrmann, G. F. (2002) Characterisation of glucose transport in *Saccharomyces cerevisiae* with plasma membrane vesicles (countertransport) and intact cells (initial uptake) with single Hxt1, Hxt2, Hxt3, Hxt4, Hxt6, Hxt7 or Gal2 transporters. *FEMS Yeast Res. 2*, 539–550.

Morgunov, I. G., and Kamzolova, S. V. Yarrowia Lipolytica Yeast Possesses An Atypical Catabolite Repression. 2011.

Ohlrogge, J. B., and Jaworski, J. G. (1997) Regulation of fatty acid synthesis. *Annu. Rev. Plant Biol. 48*, 109–136.

Pan, L. X., Yang, D. F., Li, S., Wei, L., Chen, G. G., and Liang, Z. Q. (2009) Isolation of the Oleaginous Yeasts from the Soil and Studies of Their Lipid-Producing Capacities. *Food Technology and Biotechnology 47*, 215–220.

Ratledge, C. In *Single Cell Oil*; Moreton, R., Ed.; Longman Scientific & Technical, 1988; Chapter Biochemistry, stoichiometry, substrate and economics, pp 33–70.

Ratledge, C. (2004) Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. *Biochimie 86*, 807–815.

Ratledge, C., and Wynn, J. P. (2002) The Biochemistry and Molecular Biology of Lipid Accumulation in Oleaginous Microorganisms. *Advances in Applied Microbiology 51*, 1–51.

Singh, A. (1992) Lipid accumulation by a cellulolytic strain of Aspergillus niger. *Experientia 48*, 234–6.

Walther, J. Nonstationary Metabolic Flux Analysis (NMFA) for the Elucidation of Cellular Physiology. Ph.D. thesis, Massachusetts Institute of Technology, 2010.

Wynn, J., and Ratledge, C. (1997) Malic Enzyme is a Major Source of NADPH for Lipid Accumulation by *Aspergillus Nidulans*. *Microbiology 143*, 253–257.

Wynn, J. P., bin Abdul Hamid, A., and Ratledge, C. (1999) The role of malic enzyme in the regulation of lipid accumulation in filamentous fungi. *Microbiology 145*, 1911–1917.

Wynn, J. P., Kendrick, A., and Ratledge, C. (1997) Sesamol as an inhibitor of growth and lipid metabolism in *Mucor circinelloides* via its action on malic enzyme. *Lipids 32*, 605–610.

Zhang, Y., Adams, I. P., and Ratledge, C. (2007) Malic enzyme: the controlling activity for lipid production? Overexpression of malic enzyme in Mucor circinelloides leads to a 2.5-fold increase in lipid accumulation. *Microbiology 153*, 2013–2025.

# Chapter 4

# Establishing Genetic Tools in *Yarrowia lipolytica*

## 4.1 Introduction

*Y. lipolytica* is one of the most extensively studied non-conventional yeasts. It was one of the first dozen yeast species to have their genome sequenced and over the past two decades, genetic tools and methods have been established for the engineering and manipulation of the organism. This section will discuss the genetic features and established tools for engineering *Y. lipolytica*, and the results of leveraging this knowledge for establishing and characterizing new tools for gene expression.

### 4.1.1 Transformation Vectors

The vast majority of transforming vectors in *Y. lipolytica* are shuttle vectors, which contain both yeast-derived and bacterial sequences. They typically contain a bacterial backbone for bacterial propagation consisting of a replication origin and a selection marker gene (e.g. resistance to ampicillin, kanamycin). The yeast-derived sequence contains an expression cassette (i.e. aligned promoter, gene and terminator sequences), a yeast selectable marker (e.g. leucine or uracil auxotrophy complementation, resistance to hygromycin), and possibly elements for integration and maintenance in yeast cells like docking sites and centromeres (Madzak et al., 2004).

While a survey of 24 wild-type strains of *Y. lipolytica* did not detect any endogenous plasmids (Barth and Gaillardin, 1997), replicative plasmids have been successfully used in *Y. lipolytica* (Juretzek et al., 2001; Yamane et al., 2008; Blazeck et al., 2011). Nonetheless, a majority of cloning done in *Y. lipolytica* is performed using integrative plasmids derived from a pBR322-based pINA plasmid (Juretzek et al., 2001; Madzak et al., 2004). In these cases, homologous recombination of a linearized plasmid expression cassette into a specific docking site is the method of transformation. Maximum transformation efficiencies using lithium acetate methods have been found to be on the order of $6 \times 10^5$ transformants/$\mu$g, roughly similar to that of *S. cerevisiae* (Chen et al., 1997; Gietz and Schiestl, 2007). Integration

Table 4.1: Promoters established for *Y. lipolytica* expression

| Promoter | Source | Characteristics | Reference |
|---|---|---|---|
| pLEU2 | $\beta$-isopropylmalate dehydrognase | Inducible by leucine precursor | (Madzak et al., 2000) |
| pPOX2 | Acyl-coA oxidase | Indicuble by alkanes | (Juretzek et al., 2001) |
| pXPR2 | Alkaline extracellular protease | Inducible by peptones | (Juretzek et al., 2001) |
| pICL1 | Isocitrate lyase | Inducible by acetate, ethanol, alkanes, fatty acids and derivatives | (Juretzek et al., 2001) |
| hp4d | Recombinant hybrid XPR2 promoter | Growth phase-dependent | (Madzak et al., 2000) |
| UAS1B-Leum | Recombinant hybrid XPR2 promoter | Growth phase-dependent | (Blazeck et al., 2011) |
| pTEF | Translation elongation factor-1$\alpha$ | Constitutive | (Müller et al., 1998) |
| pRPS7 | Ribosomal protein S7 | Constitutive | (Müller et al., 1998) |
| pFBA1in | Fructose-bisphosphate aldolase, with intron fusion | Constitutive | (Hong et al., 2012) |

typically results in a single copy of the expression cassette in the organism; however the use of defective selective markers has allowed multi-copy integration as extra copies of the marker are necessary to rescue the auxotrophic phenotype (Juretzek et al., 2001).

### 4.1.2 Promoters

Various promoters have been developed for expression of heterologous protein in *Y. lipolytica*. Table 4.1 summarizes most of the commonly used promoters for study and their induction conditions. The extracellular protease (XPR2) promoter was one of the first promoters characterized, allowing for strong secretion of alkaline extracellular protease up to 1-2 g/L (Madzak et al., 2004). Consequently, this promoter became the basis for numerous studies in yeast protein secretion. However, because the pXPR2 required peptone rich medium, the use of it in industrial settings was limited (Juretzek et al., 2001). Development of inducible promoters pPOX2 and pICL1 allowed for simpler expression systems that could use minimal media, while an analysis of expressed mRNA in *Y. lipolytica* for highly expressed proteins identified pTEF and pRPS7 as strongly expressing promoters (Müller et al., 1998). The

construction of hp4d was performed by arranging tandem repeats of an XPR2 upstream activator sequence (UAS) in front of a minimal LEU2 promoter. Ot was found that the UAS linearly increased expression with the number of tandem repeats (Madzak et al., 2000). The expression was found to be quasi-constitutive, with some growth phase-dependent induction in addition to a constitutive level of expression. More recently, a number of promoters have been published with improved expression above the standard expression promoters: FBA1in and UAS1B-Leum (Blazeck et al., 2011; Hong et al., 2012). These promoters have leveraged much of the knowledge base of the genomics and physiology of *Y. lipolytica* in order to induce greater expression. The FBA1in promoter utilizes introns to increase expression of the FBA promoter, while the UAS1B-Leum promoter augments the design and capacity of the hp4d promoter to increase expression.

### 4.1.3 Features of strong expression in protein encoding genes

Analysis of a number of strongly expressed genes has found common features in their sequence. The position immediately upstream of the ATG start codon is highly conserved in many of the strongly expressed genes: most predominantly CACA or CAAA. Investigation into the importance of this consensus sequence on the expression of heterologous protein human interferon alpha 2b found that the CACA sequence resulted in a 16.5-fold improvement in expression, greater than the 11-fold improvement observed through codon optimization of the gene sequence (Gasmi et al., 2011).

Introns have been detected in numerous strongly expressed genes, with about 10.6% of all genes containing at least one intron (Ivashchenko et al., 2009). The 5' end of the intron (donor site) is most commonly GTGAGT, while the 3' internal branch site has a consensus sequence of TACTAAC. The 3' terminal splice site is most commonly CAG. Intron length averages 281 bp and introns are biased towards the 5' end of the open reading frame. While the intron frequency and density varies widely among hemiascomycetous yeast, the intron sequence motifs and average lengths are well conserved, indicating mutual ancestry (Neuvéglise et al.,

2011). Spliceosomal introns have been implicated in improving expression in a variety of yeasts, including *Y. lipolytica* (Le Hir et al., 2003; Hong et al., 2012).

### 4.1.4 Establishing genetic tools for *Y. lipolytica*

Well known to be a strong constitutively expressed gene, translation elongation factor 1-$\alpha$ (TEF) contains all the features of strong expression. Both the CAAA sequence and the spliceosomal intron can be found in the sequence of TEF, which is shown in Figure 4.1.1. Indeed, TEF was originally identified by its enrichment in cDNA libraries (Müller et al., 1998). As such, the TEF promoter was used as the basis for the construction of an expression platform in *Y. lipolytica*. Additionally, the signifiance and effects of the spliceosomal intron in the vector are characterized. Furthermore we will describe and validate preliminary work on establishing an alternative expression vector, capable of complementing gene expression of the pIN1269 vector. From the existing knowledge base, we establish additional genetic tools for *Y. lipolytica* and an expression vector that can be used in future studies for the metabolic engineering of *Y. lipolytica*.

## 4.2 Materials & Methods

### 4.2.1 Yeast strains, growth, and culture conditions

The *Y. lipolytica* strains used in this study were derived from the wild-type *Y. lipolytica* W29 strain (ATCC20460). The auxotrophic Po1g (Leu-) used in all transformations was obtained from Yeastern Biotech Company (Taipei, Taiwan). All strains used in this study are listed in Table 4.2.

Media and growth conditions for *Escherichia coli* have been previously described by Sambrook et al. (2001), and those for *Y. lipolytica* have been described by Barth and Gaillardin (Barth and Gaillardin, 1997). Rich medium (YPD) was prepared with 20 g/L Bacto peptone (Difco Laboratories, Detroit, MI), 10 g/L yeast extract (Difco), 20 g/L glucose

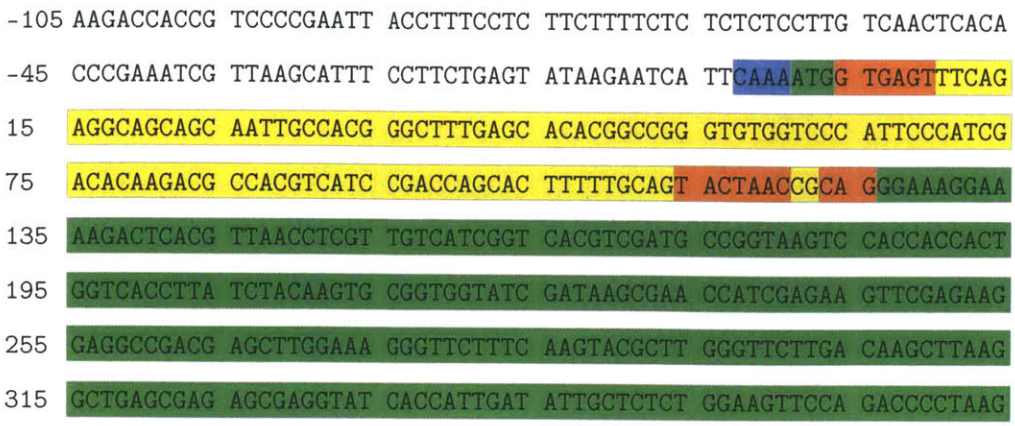


Figure 4.1.1: Genetic features of expression in Translation Elongation Factor-1$\alpha$ (TEF) both upstream and downstream of the start codon. CACA or CAAA sequences (-4 to -1; highlighted in blue) are highly conserved immediately upstream of the start codon. A spliceosomal intron (4 to 125; highlighted in yellow) is found in highly expressed open reading frames, and exhibit expression enhancing characteristics. In the TEF open reading frame (highlighted in green), the spliceosomal intron is immediately after the start codon. The functional consensus sequences for the 5' splice site, the branch site and the 3' splice site are denoted in orange.

(Sigma-Aldrich, St. Louis, MO). YNB medium was made with 1.7 g/L yeast nitrogen base (without amino acids) (Difco), 0.69 g/L CSM-Leu (MP Biomedicals, Solon, OH), and 20 g/L glucose. Selective YNB plates contained 1.7 g/L yeast nitrogen base (without amino acids), 0.69 g/L CSM-Leu, 20 g/L glucose, and 15 g/L Bacto agar (Difco).

Shake flask experiments were carried out using the following medium: 1.7 g/L yeast nitrogen base (without amino acids), 1.5 g/L yeast extract, and 50 g/L glucose. From frozen stocks, precultures were inoculated into YNB medium (5 mL in Falcon tube, 200 rpm, 28°C, 24 hr). Overnight cultures were inoculated into 50 mL of media in 250 mL Erlenmeyer shake flask to an optical density (A600) of 0.05 and allowed to incubate for 100 hours (200 rpm, 28°C), after which biomass, sugar content, and lipid content were taken and analyzed.

### 4.2.2 LacZ Plasmid Construction

Standard molecular genetic techniques were used throughout this study (Sambrook and Russell, 2001). Restriction enzymes and Phusion High-Fidelity DNA polymerase used in

Table 4.2: Strains and plasmids used in this study

| **Strains (host strain)** | **Genotype or plasmid (Selective Marker)** | **Source** |
|---|---|---|
| *E. coli* | | |
| DH5$\alpha$ | fhuA2 $\Delta$(argF-lacZ)U169 phoA glnV44 $\Phi$80 $\Delta$(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 | Invitrogen |
| pINA1269 | JMP62-LEU2 | Yeastern |
| JMP62-URA3 | JMP62-URA3 | Nicaud et al. |
| pDUET-1 | pACYCDUET-1 | Novagen |
| pMT010 | pINA1269 php4d::TEF (LEU) | This work |
| pMT015 | pINA1269 php4d::TEFin (LEU) | This work |
| pMT025 | hp4d-LacZ (LEU) | This work |
| pMT038 | YTEF-LacZ (LEU) | This work |
| pMT037 | YTEF(intron)-LacZ (LEU) | This work |
| pMT043 | hp4d(intron)-LacZ (LEU) | This work |
| pMT053 | TEFin-DGA (LEU) | Chapter 6 |
| pMT091 | pACYC Lip2 KO plasmid (URA) | This work |
| pMT092 | pMT091 + TEFin-DGA (URA) | Chapter 7 |
| *Y. lipolytica* | | |
| Po1g | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 | Yeastern |
| Po1z | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 $\triangle$URA3 | This work |
| MTYL025 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 hp4d-LacZ | This work |
| MTYL037 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEFin-LacZ-LEU2 | This work |
| MTYL038 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-LacZ-LEU2 | This work |
| MTYL043 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 hp4din-LacZ-LEU2 | This work |
| MTYL053 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEFin-DGA | Chapter 6 |
| MTYL092 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 LIP2::TEFin-DGA | Chapter 7 |

cloning were obtained from New England Biolabs (Ipswich, MA). Genomic DNA from yeast transformants was prepared using Yeastar Genomic DNA kit (Zymo Research, Irvine, CA). All constructed plasmids were verified by sequencing. PCR products and DNA fragments were purified with PCR Purification Kit or QIAEX II kit (Qiagen, Valencia, CA). Plasmids used are described in Table 4.2. Primers used are described in Table 4.3.

Plasmid pMT010 was constructed by amplifying the translation elongation factor-1$\alpha$ (TEF) promoter region (Accession number: AF054508) from *Y. lipolytica* Po1g genomic DNA using primers MT078 and MT079. The amplicon was inserted between SalI and KpnI sites of the starting vector, pINA1269, also known as pYLEX1, obtained from Yeastern Biotech Company (Taipei, Taiwan). Also included in the reverse primer MT079 were MluI and NsiI sites to add restriction sites to the multi-cloning site.

Plasmid pMT015 was constructed by amplifying from *Y. lipolytica* Po1g genomic DNA the TEF promoter and the 5' coding region containing the ATG start codon and 113 bp of the endogenous intron (Accession number: CR382129). Primers MT118 and MT122 were used for this amplification and inserted between SalI and MluI sites of pMT010. For cloning purposes, some of the intron was omitted so that the SnaBI restriction site could be incorporated. Cloning a gene into this plasmid thus requires the omission of the gene's ATG start codon, addition of TAACCGCAG to the beginning of the 5' primer, and blunt-end ligation at the 5' end. The plasmid map is shown in Figure 4.2.1.

Plasmid pMT025 was constructed by amplifying the LacZ gene, encoding $\beta$-galactosidase, from *E. coli* (K12) and inserting it into the PmlI and BamHI sites of starting vector pINA1269 using primers MT170 and MT171. Plasmid pMT038 was constructed by amplifying the LacZ gene and inserting it into the MluI and NsiI sites of pMT010 using primers MT168 and MT169. Since LacZ contains multiple MluI sites, AscI was used as the 5' restriction site on MT168 that has a matching overhang. Plasmid pMT037 was constructed by amplifying LacZ gene and inserting it into the SnaBI and NsiI sites of pMT015. Primers MT172 and MT169 were used, where forward primer MT127 omits the ATG start codon of LacZ and instead

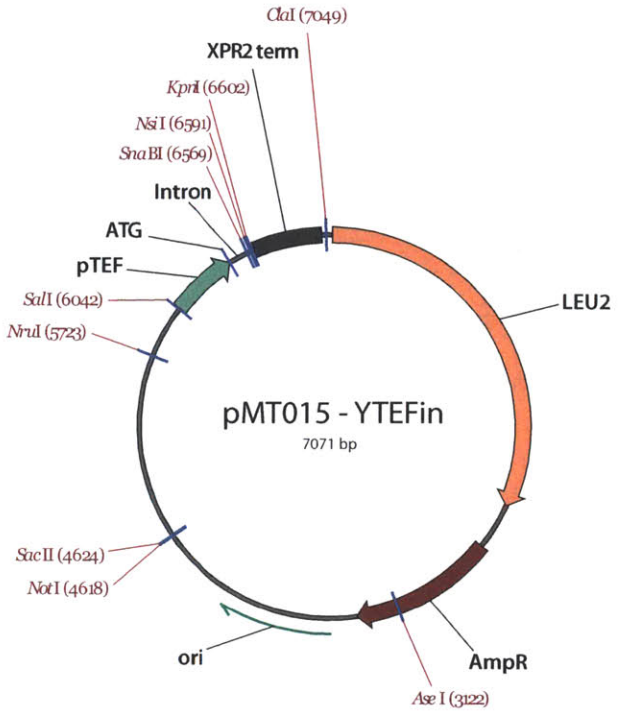


Figure 4.2.1: Plasmid map of pMT015, YTEFin, for cloning under the control of intronic TEF promoter. Cloning site exists between SnaBI and KpnI.

begins with the sequence TAACCGCAG that completes the intron sequence of pMT015. Plasmid pMT043 was constructed by amplifying from constructed plasmid pMT037 using primer pair MT211 and MT171. This was then inserted into plasmid pINA1269 using restriction sites PmlI and BamHI.

### 4.2.3 $\beta$-galactosidase assay

LacZ enzyme activity was measured using the $\beta$-gal assay kit from Sigma-Aldrich. Cells were resuspended in Phosphate Buffered Saline (PBS) buffer and lysed by vortexing with 500 $\mu$m glass beads (Sigma-Aldrich) for 2 minutes. 40 $\mu$L of the cell lysate was transferred into 340 $\mu$L reaction mix containing 47 mg/mL ONPG, 0.6 M $Na_2HPO_4{\cdot}7H2O$, 0.4 M $NaH_2PO_4{\cdot}H_2O$, 0.1 M KCl, 0.01 M $MgSO_4{\cdot}7H_2O$. Reaction was incubated at 37°C for color evolution to occur, and was finally quenched using 500 $\mu$L 1 M Sodium Carbonate. Absorbance was then

Table 4.3: Primers used in this study. Relevant restriction sites are in bold.

| Primer | Description | Sequence |
|---|---|---|
| *PCR* | | |
| MT078 | TEF | *GACT**GTCGAC**AGAGACCGGGTTGGCGGCGCATTTGTG* |
| MT079 | TEF | *GACTGGTACCTCAAGATGCATAGCACGCGTTTTGAATGATTCTTATACTC* |
| MT118 | TEFin | *GGCA**GTCGAC**AGAGACCGGGTTGGCGGC* |
| MT122 | TEFin | *TTATTCACGCGTGTAGATACGTACTGCAAAAAGTGCTGGTCGGA* |
| MT170 | LacZ | *AATGACCATGATTACGGATTCACTGG* |
| MT171 | LacZ | *CTAGGT**GGATCC**TTATTTTTGACACCAGACCAACTGGTAA* |
| MT172 | LacZ | *TAACCGCAGACCATGATTACGGATTCACTGGCC* |
| MT173 | LacZ | *CTAGGT**ATGCAT**ATGACCATGATTACGGATTCACTGG* |
| MT174 | LacZ | *CTTACA**GGTACC**TTATTTTTGACACCAGACCAACTGGTAA* |
| MT211 | LacZ | *AATGGTGAGTTTCAGAGGCAGCAG* |
| MT310 | *Yl*URA UP | *CTCAAGCTCGTGGCAGCCAA* |
| MT311 | *Yl*URA UP | *GGCAATGAAGCCTGGTGCTTGACAGTGTTGCCA* |
| MT312 | *Yl*URA DOWN | *GTCAAGCACCAGGCTTCATTGCCCAGAACCGAC* |
| MT313 | *Yl*URA DOWN | *GGTATCGCTTGGCCTCCTCAAT* |
| MT314 | URA3 | *CTCACT**CGATCG**TATCGATCCGAGAAACACAACAACATGCC* |
| MT315 | URA3 | *CTCACT**GGTACC**GCCCAGAGAGCCATTGACGTTC* |
| MT316 | Lip2 UP | *CTCACT**GGATCC**CCGCGGCCACCATCCTCTTCACAGCCTG* |
| MT317 | Lip2 UP | *CTCACT**GAATTC**TCGTCAGAGGAGCCTGCATGAT* |
| MT318 | Lip2 DOWN | *CTCACT**GGTACC**CCAGATTGCTGTCACCGGTCA* |
| MT319 | Lip2 DOWN | *CTCACT**CCTAGG**CCGCGGCCCTCGGTGACGAAGTACTGCA* |

measured with a spectrophotometer at 420 nm. Enzymatic units are calculated based on enzyme activity divided by incubation time and dry cell weight.

### 4.2.4 Deletion of URA3

To increase the availability of markers in strains of *Y. lipolytica*, the gene encoding for uracil prototrophy, orotidine-5'-phosphate decarboxylase (URA3, Accession Number: AJ306421), was amplified and used as the basis of a knockout cassette for the generation of URA auxotrophic strains. Upstream and downstream sequences of the URA open reading frame were amplified using primer pairs MT310 - MT311 and MT312 - MT313, respectively. The primers are designed such that the two amplicons carry 23 bp overlapping region. Upon purification of the two amplicons, both products are mixed and a PCR is performed using the primers MT310 and MT313 to produce a 456 bp amplicon fusing the upstream and downstream amplicons. This DNA was purified and subsequently transformed into Po1g. Transformed cells were then plated on a selective media plate containing uracil and 5-Fluoroorotic Acid

(5-FOA). Colonies that grew were replated on 5-FOA plates to reselect for URA auxotrophy, and verified by PCR of prepared genomic DNA. The resulting ΔLEU2 ΔURA3 strain was named Po1z.

### 4.2.5 Complementary vector construction

For the construction of a complementary vector, pACYCDUET-1 was selected as the complementary shuttle vector, as it utilizes a different backbone, selective marker and origin of replication. The upstream sequences of LIP2 (Accession Number: XM_500282) were amplified from *Y. lipolytica* Po1g genomic DNA using the primer pairs MT316 - MT317 and integrated into pDUET using the restriction sites BamHI and EcoRI. The downstream sequence of LIP2 was amplified using primer pairs MT318 - MT319 and integrated into pDUET using the restriction sites KpnI and AvrII. The selective marker for yeast uracil prototrophy was amplified from the plasmid JMP62-URA using primers pairs MT314 - MT315 and integrated into pDUET using the restriction sites PvuI and KpnI. The resulting plasmid pMT091 contained a multi-cloning site flanked by upstream and downstream LIP2 sequences and a URA3 marker. Digestion with the restriction enzyme SacII linearizes the plasmid and separates the integration vector from the plasmid backbone, minimizing the integration of foreign and unnecessary DNA. Use of the complementary plasmid requires construction of the expression cassette on the original pINA1269 backbone and then transferring the cassette over to pMT091 using restriction enzyme subcloning. The large multi-cloning site and differential antibiotic marker facilitate the cloning and selection process. The plasmid map for pMT091 is depicted in Figure 4.2.2.

The cloning of pMT092 is detailed in Chapter 7 on page 119. Briefly, the expression cassette of TEFin-DGA was cut from plasmid pMT053 using restriction sites SalI and EcoRI. This fragment was ligated into pMT091 using the same restriction sites. For transformation, the plasmid pMT092 was digested with restriction enzyme SacII and transformed into Po1z, the ΔURA3 mutant of Po1g. The strain was verified by PCR of prepared genomic DNA.

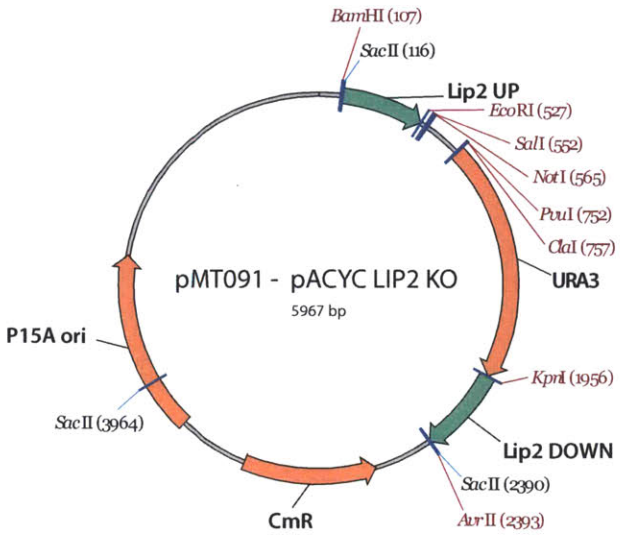


Figure 4.2.2: Plasmid map of pMT091, for simultaneous knockout of LIP2 and integration of expression cassette. Multi-cloning site exists between EcoRI and ClaI.

### 4.2.6 RNA isolation and transcript quantification

Shake flask cultures grown for 42 hrs were collected and centrifuged for 5 min at 10,000g. Each pellet was resuspended in 1.0 ml of Trizol reagent (Invitrogen) and 100 $\mu$L of acid-washed glass beads were added (Sigma-Aldrich). Tubes were vortexed for 15 min at 4°C for cell lysis to occur. The tubes were then centrifuged for 10 min at 12,000g at 4°C and the supernatant was collected in a fresh 2-mL tube. 200 $\mu$L chloroform was then added and tubes were shaken by hand for 10 seconds. The tubes were again centrifuged for 10 min at 12,000g at 4°C. 400 $\mu$L of the upper aqueous phase was transferred to a new tube, and an equal volume of phenol-chloroform-isoamyl alcohol (pH 4.7) (Ambion, Austin, TX) was added. Tubes were again shaken by hand for 10 seconds and centrifuged for 10 min at 12,000g at 4°C. 250 $\mu$L of the upper phase was transferred to a new tube with an equal volume of cold ethanol and 1/10th volume sodium acetate (pH 5.2). Tubes were chilled at -20°C for thirty minutes to promote precipitation. Tubes were then centrifuged for 5 min at 12,000g, washed twice with 70% ethanol, dried in a 60°C oven and finally resuspended in RNAse free water. RNA quantity was analyzed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and samples were stored in -80°C freezer. qRT-PCR analyses


Figure 4.3.1: Assay for LacZ activity. Eight samples expressing LacZ from *E. coli* after cell lysis and incubation for 30 minutes with X-gal in 1x PBS. Samples A-H are biological replicates of strain MTYL025.

were carried out using iScript One-step RT-PCR Kit with SYBR Green (Bio-Rad, Hercules, CA) using the Bio-Rad iCycler iQ Real-Time PCR Detection System. Fluorescence results were analyzed using Real-time PCR Miner and relative quantification and statistical analysis was determined with REST 2009 (Qiagen) using actin as the reference gene and MTYL038 as the reference strain (Zhao and Fernald, 2005). Samples were analyzed in quadruplicate.

## 4.3 Results & Discussion

### 4.3.1 Selection of reporter gene

In order to quantify the strength of expression of our expression vector, it is useful to use a reporter gene to help with characterization. The most commonly used reporter system is green fluorescence protein (GFP) and its variants, all of which have been used extensively to characterize everything from expression vectors to entire promoter libraries, in a wide range of organisms from bacteria to mammals (Alper et al., 2005; Zimmer, 2002). However, it is peculiar that until recently, there have been relatively few reported literature citations on the use of GFP in *Y. lipolytica*, and typically only in cases of fusion proteins (Ruiz-Pavón and

Domínguez, 2007; Yue et al., 2008). Our failures in experimentation seem to corroborate the lack of evidence in literature: we were unable to functionally express GFP in *Y. lipolytica.* Using different promoters (pPOX2, pTEF, pRPS7), vectors (pINA1269, JMP62), or GFP variants (GFP, yEGFP, superGFP) did not seem to have any observable effect (data not shown)[1].

Instead of fluorescence, enzymatic assays have typically been employed to quantify expression in *Y. lipolytica.* The enzyme $\beta$-galactosidase is a very common enzyme that can be used to expression both qualitatively and quantitatively. Expression of bacterial LacZ in *Y. lipolytica* successfully resulted in detectable $\beta$-galactosidase activity. It was necessary to lyse the cells through agitation with glass beads, as virtually no protein secretion of the enzyme was detected (in contrast with bacterial expression of the LacZ gene). Qualitative results of the activity of $\beta$-galactosidase on the reporter molecule X-gal is depicted in Figure 4.3.1. For quantification of enzyme activity, an ortho-Nitrophenyl-$\beta$-galactoside (ONPG) assay was used with a UV/Vis spectrophotometer.

### 4.3.2 A strong expression platform based on TEF is enhanced by spliceosomal intron

In *Y. lipolytica*, several promoters are available for gene expression, including inducible and constitutive ones (Madzak et al., 2004). The TEF promoter was originally identified as being a strong constitutive promoter; however, subsequent cloning, characterization and other alterations resulted in lower expression relative to the inducible XPR2 promoter (Müller et al., 1998). More recently, the hybrid hp4d promoter has been used for its strong quasi-constitutive expression (Madzak et al., 2000), and has been employed in a number of appli-

[1]It was not until recently that Blazeck et al. published results showing that while most GFP variants did not work (i.e. yECitrine, mStrawberry, EGFP), human variant GFP could properly be expressed and fluoresce(Blazeck et al., 2011). They reasoned that the codon usage in the hrGFP was the only protein sequence similar enough to *Y. lipolytica* that proper expression could be achieved.

cations requiring high protein expression (Chuang et al., 2010; Cui et al., 2011; Gasmi et al., 2011).

Analysis of the TEF genomic sequence reveals the presence of a 122-bp spliceosomal intron immediately after the start codon in the 5' region of the open reading frame (Figure 4.1.1). Promoter-proximal spliceosomal introns have often been found to dramatically affect expression of their corresponding genes in a variety of organisms (Le Hir et al., 2003). We hypothesized that the intron impacted strong expression of TEF, and that stronger expression could be achieved by including the intron along with the promoter in the expression vector. Indeed, the initial screening and isolation of the TEF promoter likely relied on the intron-enhanced enrichment in cDNA libraries, a feature that would not have been noticed once the intron was spliced.

To further investigate the effect of introns and compare their impact on relative gene expression driven by three promoters, plasmids pMT025, pMT037 and pMT038 were constructed expressing $\beta$-galactosidase (LacZ) to compare the relative expression of three promoters (Figure 4.3.2): synthetic hybrid promoter (php4d), TEF promoter without intron (pTEF), and TEF promoter with intron (pTEFintron). Remarkably, the TEFin promoter exhibited a 12-fold increase in expression over the intronless TEF promoter, and a 5-fold increase in expression over the hp4d promoter after 50 hrs of culture.

Time course experiments of the promoters showed that TEF(intron) expression peaks at about 50 hrs, likely related to the activity of the spliceosomal machinery in the cell during different growth phases. These results are depicted in Figure 4.3.3. The hp4d promoter exhibits stronger expression in the early growth phase (before 25 hrs), consistent with the quasi-growth dependent characterization. Expression from a hybrid hp4d(intron) promoter did not exhibit any improved expression over hp4d. This may be an indication of specialized relationships between the promoter and intron sequences that are not present in the truncated, minimal hybrid promoter. As such, it remains to be seen whether the expression enhancing characteristics of the intron are interchangeable with other endogenous promot-

ers. Other work in *Y. lipolytica* on expression enhancing introns seems to indicate in the affirmative (Hong et al., 2012).

The intron enhancement observed in other systems varies wildly: from only 2-fold in human cells and yeast to over 1000-fold in maize (Callis et al., 1987; Furger and Binnie, 2002). Introns are believed to enhance gene expression in a number of ways: by containing regulatory elements, facilitating mRNA export, and increasing transcription initiation rates (Le Hir et al., 2003). Intronic genes, as a group, tend to exhibit higher levels of expression relative to non-intronic genes. For example, in *S. cerevisiae*, intronic genes only represent less than 4% of the total gene count, yet account for 27% of the total RNA generated in the cell (Ares et al., 1999). The genome of *Y. lipolytica* contains introns in 10.6% of its genes, compared to only 4.5% in *S. cerevisiae* (Ivashchenko et al., 2009). Enlisting this endogenous process to enhance expression of our own desired genes represents a simple means for modulating pathway flux, applicable to a broad range of eukaryotic organisms. The high sequence homology of splice sequences among hemiascomycetous yeast, indicates that these introns may be cross-compatible (Bon et al., 2003). While research continues to further elucidate the function, evolution, and purpose of introns, the utilization of introns for biotechnological purposes is a relatively untapped opportunity.

### 4.3.3 Complementary vector construction allows for alternative method for expression of DGA

In order to provide an alternative and complementary method for expressing genes in *Y. lipolytica*, the URA marker was knocked out of the parent strain Po1g. As part of the design of a complementary integration vector, the extracellular lipase LIP2 gene was selected as the docking site, as this gene is well characterized but likely has a negligible or neutral effect on *de novo* lipid synthesis and accumulation. Thus in the process of integrating the complementary vector, LIP2 will be knocked out. Upon construction of the complementary vector for gene expression, it was necessary to examine whether the expression of a gene

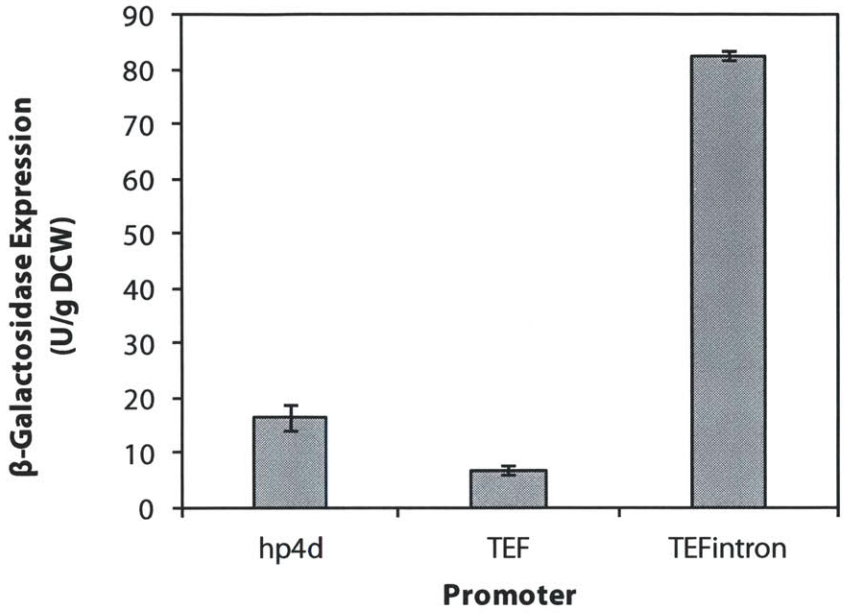


Figure 4.3.2: Enzyme activity of $\beta$-galactosidase (LacZ) under the control of different promoters after 50 hours of culture. Samples performed in duplicate.

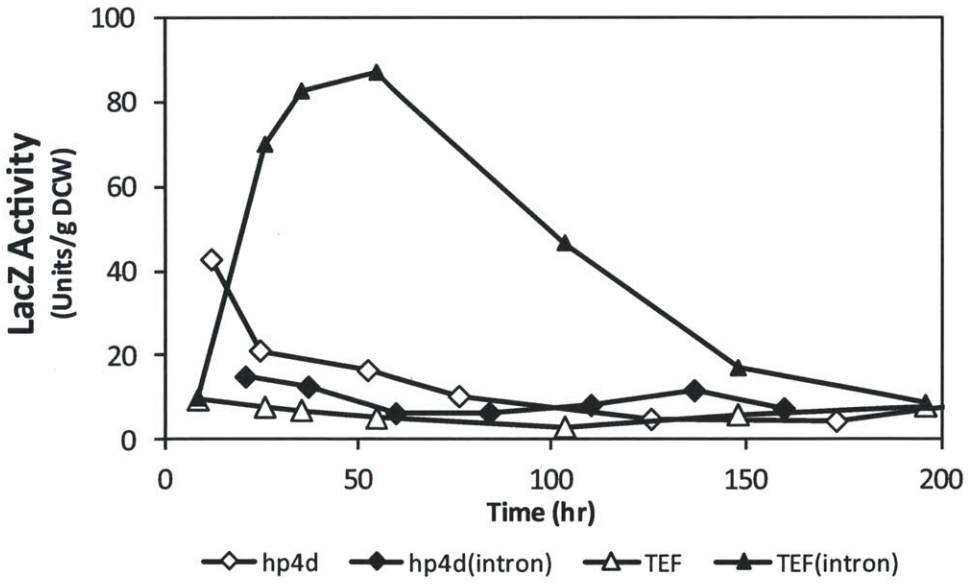


Figure 4.3.3: LacZ expression over time under TEF/hp4d promoters with or without intron

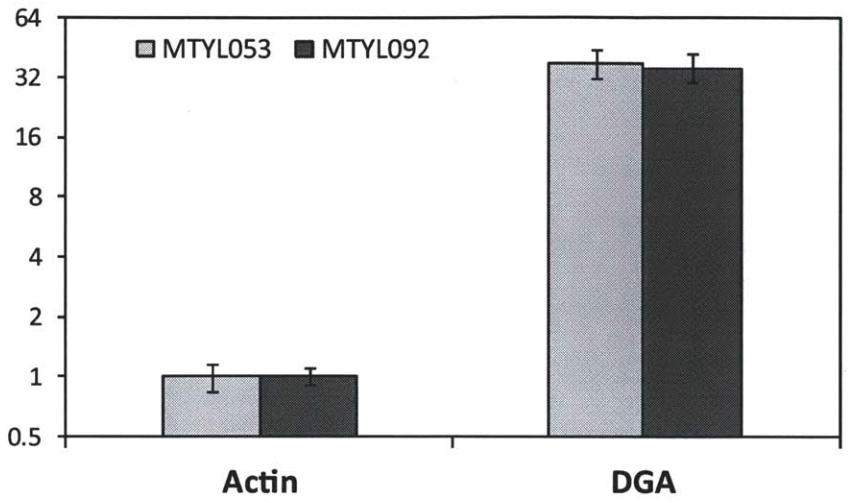


Figure 4.3.4: Transcriptional expression of TEFin-DGA cassette on original plasmid (PMT053) vs. complementary plasmid (PMT092).

would be affected depending on which transformation vector was used. To test this, the gene encoding for diacylglycerol acyltransferase (DGA) was cloned into both pMT015 and pMT091 vectors. While the expression cassette was the same in both vectors, the docking site was different: a pBR322 docking site for pMT015 (and all pINA1269-based vectors), and the LIP2 gene for pMT091. Upon transformation of these vectors and verification of genomic integration, RT-PCR of extracted RNA was performed on both strains, examining the overexpression of DGA in both cases relative to a control strain (MTYL038). As shown in Figure 4.3.4, both strains exhibit 32-fold increases in DGA expression relative to MTYL038 control strains. Lipid accumulation due to DGA overexpression was also similar (data not shown). These results validate the use of pMT091 as a complementary vector and LIP2 as a alternative docking site for gene expression, with no detectable interference from possible epigenetic phenomena.

## 4.4 Conclusion

Expanding the genetic toolbox will continue to be necessary for efficient and robust engineering of the non-conventional yeast *Y. lipolytica*. Utilization of the 5' spliceosomal intron

from the TEF open reading frame was found to dramatically increase expression. This helps us establish a strong expression platform to overexpress potential genes of interest and lends support to the utilization of expression-enhancing introns for metabolic engineering purposes. Additionally, the construction of a secondary plasmid that can be used to further increase the integration of expression cassettes allow for the study of the interactions of numerous genes in cooperation. Through these tools, we can utilize strategies of metabolic engineering to investigate and engineer *Y. lipolytica* for the production of biofuels.

# References


Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A. 102*, 12678–12683.

Ares, M., Jr, Grate, L., and Pauling, M. H. (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA 5*, 1138–1139.

Barth, G., and Gaillardin, C. (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica. FEMS Microbiol. Rev. 19*, 219–237.

Blazeck, J., Liu, L., Redden, H., and Alper, H. (2011) Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. *Appl. Environ. Microbiol. 77*, 7905–7914.

Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuvéglise, C., Munsterkotter, M., Guldener, U., Mewes, H. W., Van Helden, J., Dujon, B., and Gaillardin, C. (2003) Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res. 31*, 1121–1135.

Callis, J., Fromm, M., and Walbot, V. (1987) Introns increase gene expression in cultured maize cells. *Genes & Development 1*, 1183–1200.

Chen, D. C., Beckerich, J. M., and Gaillardin, C. (1997) One-step transformation of the dimorphic yeast *Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 48*, 232–235.

Chuang, L.-T., Chen, D.-C., Nicaud, J.-M., Madzak, C., Chen, Y.-H., and Huang, Y.-S. (2010) Co-expression of heterologous desaturase genes in *Yarrowia lipolytica. New Biotechnology 27*, 277–282.

Cui, W., Wang, Q., Zhang, F., Zhang, S.-C., Chi, Z.-M., and Madzak, C. (2011) Direct conversion of inulin into single cell protein by the engineered *Yarrowia lipolytica* carrying inulinase gene. *Process Biochemistry 46*, 1442–1448.

Furger, A., and Binnie, J. (2002) Promoter proximal splice sites enhance transcription. *Genes & Development 16*, 2792.

Gasmi, N., Ayed, A., Nicaud, J. M., and Kallel, H. (2011) Design of an efficient medium for heterologous protein production in *Yarrowia lipolytica*: case of human interferon alpha 2b. *Microbial Cell Factories 10*, 38.

Gietz, R. D., and Schiestl, R. H. (2007) High-efficiency yeast transformation using the Li-Ac/SS carrier DNA/PEG method. *Nature Protocols 2*, 31–34.

Hong, S.-P., Seip, J., Walters-Pollak, D., Rupert, R., Jackson, R., Xue, Z., and Zhu, Q. (2012) Engineering *Yarrowia lipolytica* to express secretory invertase with strong FBA1IN promoter. *Yeast 29*, 59–72.

Ivashchenko, A. T., Tauasarova, M. I., and Atambayeva, S. A. (2009) Exon-intron structure of genes in complete fungal genomes. *Molecular Biology 43*, 24–31.

Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G., and Nicaud, J. (2001) Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast 18*, 97–113.

Le Hir, H., Nott, A., and Moore, M. J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci. 28*, 215–220.

Madzak, C., Gaillardin, C., and Beckerich, J. M. (2004) Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *J. Biotechnol. 109*, 63–81.

Madzak, C., Tréton, B., and Blanchin-Roland, S. (2000) Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *J. Mol. Microbiol. Biotechnol. 2*, 207–216.

Müller, S., Sandal, T., Kamp-Hansen, P., and Dalbøge, H. (1998) Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. *Yeast 14*, 1267–1283.

Neuvéglise, C., Marck, C., and Gaillardin, C. (2011) The intronome of budding yeasts. *Comptes Rendus Biologies 334*, 662–670.

Nicaud, J.-M., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P., and Gaillardin, C. (2002) Protein expression and secretion in the yeast *Yarrowia lipolytica*. *FEMS Yeast Res. 2*, 371–379.

Ruiz-Pavón, L., and Domínguez, A. (2007) Characterization of the Yarrowia lipolytica Yl-SRP72 gene, a component of the yeast signal recognition particle. *Int Microbiol 10*, 283–289.

Sambrook, J., and Russell, D. W. *Molecular cloning: a laboratory manual*; CSHL press, 2001; Vol. 2.

Yamane, T., Sakai, H., Nagahama, K., Ogawa, T., and Matsuoka, M. (2008) Dissection of centromeric DNA from yeast *Yarrowia lipolytica* and identification of protein-binding site required for plasmid transmission. *J. Biosci. Bioeng. 105*, 571–578.

Yue, L., Chi, Z., Wang, L., Liu, J., Madzak, C., Li, J., and Wang, X. (2008) Construction of a new plasmid for surface display on cells of *Yarrowia lipolytica*. *Journal of Microbiological Methods 72*, 116–123.

Zhao, S., and Fernald, R. D. (2005) Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. *J. Comput. Biol. 12*, 1047–1064.

Zimmer, M. (2002) Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chem Rev 102*, 759–781.

# Chapter 5

# Intron Bioinformatics Elucidates Yeast Intron Evolution

In addition to establishing genetic tools for *Y. lipolytica*, particularly the implementation of strong expression through the use of spliceosomal introns, it was instructive to explore the broader evolutionary context of these introns. This section discusses the use of comparative bioinformatic analysis to look at the evolution of introns in yeast species.

## 5.1 Background

Through the use of modern bioinformatic tools, large scale genome sequencing and annotation has enabled broad analyses of the relationships and orthology of genetic information across different organisms. In particular, work on comparative genomics of related species continues to highlight the importance of conserved sequences and genetic elements, elucidating both functional and evolutionary relationships. In light of this, the comparison of genomes across yeast species offers an excellent opportunity to investigate the selective pressures that influence intron evolution and their possible role in yeast evolution (Dujon, 2006). What has been found is that most yeast intronic genes are limited to a singular spliceosomal intron near the 5' region. Compared to higher eukaryotic organisms, where multiple introns per gene permit alternative splicing activity, yeast introns would thus seem to have limited functionality. The preference for the 5' end is thought to be due to the mechanism for intron removal: homologous recombination of spliced cDNA beginning at the 3' end (Fink, 1987). Yeast are also unique in that evolutionarily they appear to be undergoing a process of intron loss, exhibited by a transition from intron-rich to intron-poor genomes (Dujon, 2006). Because of this transition, intron densities vary dramatically: from greater than 95% of genes in *Cryptococcus neoformans* to less than 1% in *Encephalitozoon cuniculi* (Ivashchenko et al., 2009). The loss of yeast introns would indicate that introns impose a fitness penalty to these organisms and thus their removal confers a competitive advantage. Indeed, evolution of unicellular organisms is often characterized by intense optimization towards higher growth

rates (Jeffares et al., 2006). Even the slightest advantages would allow one organism in a dense, fast-growing community to overtake the population.

It has also been observed that the use of the spliceosome significantly extends the time for gene expression for a given transcript. Transcription typically proceeds at the rate of 1 kb per minute while intron excision via the spliceosome can take up to 3 minutes (Neugebauer, 2002). For yeast with genes averaging 1.6 kb in size (Lewin, 2008), intron removal represents a significant time commitment. One theory of intron loss suggests that this time delay imposes a substantial fitness penalty on microorganisms, and thus gives rise to selective pressures for their removal (Jeffares et al., 2006). However, despite these selective pressures, introns have not been completely eliminated from yeast genomes; indeed, some introns are even necessary for proper cell growth and viability (Parenteau et al., 2008). Interestingly, many of these spliceosomal introns have been observed and identified to increase the expression of their associated gene, likely due to positive interactions with the spliceosomal machinery (Le Hir et al., 2003). In maize, where this phenomenon was first studied, introns affected expression of alcohol dehydrogenase-1 (ADH1) between 50- to 100-fold (Callis et al., 1987). In transgenic mice, a recombinant intron was used to increase gene expression up to 5- to 300-fold (Choi et al., 1991). In some cases, perhaps this increased expression, depending on function of the associated gene, may confer a fitness advantage in spite of the lengthier expression process. This trade-off may play a significant role in yeast evolution as the organism attempts to balance the costs and benefits of utilizing introns.

Organisms in dense populations experience intense competition for resources and strong selection for rapid growth (Jeffares et al., 2006). We hypothesize that as yeast species evolve in such an environment, only the most critical genes providing a net fitness benefit from intron expression enhancement will remain; enrichment of introns in particular pathways is a consequence of optimization towards higher fitness. Identifying these relationships can thus allow us to better understand the cellular pathways important to yeast evolution as well as provide insight into species-specific adaptations to their respective environments.

Table 5.1: Abbreviations of yeast studied and calculated intronic gene densities

| Abbreviation | Organism Name | Intronic Gene Density |
|---|---|---|
| **AGO** | *Ashbya gossypii* | 7.0% |
| **CTP** | *Candida tropicalis* | 1.7% |
| **CGR** | *Candida glabrata* | 2.7% |
| **DHA** | *Debaryomyces hansenii* | 10.8% |
| **KLA** | *Kluyveromyces lactis* | 4.3% |
| **NCR** | *Neurospora crassa* | 88.9% |
| **PPA** | *Pichia pastoris* | 13.5% |
| **PIC** | *Scheffersomyces/Pichia stipitis* | 21.4% |
| **SCE** | *Saccharomyces cereivisiae* | 5.9% |
| **SPO** | *Schizosaccharomyces pombe* | 47.4% |
| **YLI** | *Yarrowia lipolytica* | 24.5% |
| **ZRO** | *Zygosaccharomyces rouxii* | 4.2% |

## 5.2 Functional Intron Distributions in Yeast

In order to test this hypothesis, we analyzed intronic gene/protein distributions in twelve ascomycetous fungal genomes, predominantly in the hemiascomycetous phylum, and compared them to gene-function information according to the KEGG Orthology (KO) database. We then performed over-representation analysis of the intronic gene distributions to determine whether any cellular pathways were enriched above statistical baseline. For each pathway category, an enrichment score was calculated as the negative log of the p-value for over-representation, which was calculated using the Fisher's exact test based on a hypergeometric distribution (Huang et al., 2009). The scores for each pathway were plotted and compared to each organism in accordance to their genomic phylogeny to approximate the path of evolution in yeast species (Figure 5.2.1). Abbreviations and intronic gene density data can be found in Table 5.1. Intron frequency data used to calculate these enrichment scores can be found in Tables 5.2 and 5.3. The underlying MATLAB code for this procedure can be found in Appendix A.

Table 5.2: Intronic gene frequencies according to KEGG Pathways. *Genes* column denotes the total number of genes contained to the pathway. *Introns* column denotes the number of identified intronic genes contained in the pathway.

| **Cellular Pathway** | **AGO** | | **CTP** | | **CGR** | | **DHA** | | **KLA** | | **NCR** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Genes* | *Introns* | *Genes* | *Introns* | *Genes* | *Introns* | *Genes* | *Introns* | *Genes* | *Introns* | *Genes* | *Introns* |
| Carbohydrate Metabolism | 222 | 3 | 200 | 3 | 268 | 0 | 160 | 6 | 235 | 0 | 278 | 257 |
| Energy Metabolism | 90 | 7 | 87 | 4 | 93 | 3 | 71 | 10 | 95 | 4 | 126 | 115 |
| Lipid Metabolism | 96 | 3 | 78 | 0 | 106 | 0 | 64 | 1 | 98 | 2 | 113 | 100 |
| Nucleotide Metabolism | 135 | 5 | 118 | 0 | 138 | 4 | 66 | 5 | 141 | 3 | 130 | 111 |
| Amino Acid Metabolism | 202 | 3 | 176 | 0 | 222 | 0 | 146 | 7 | 214 | 0 | 273 | 244 |
| Metabolism of Other Amino Acids | 42 | 1 | 53 | 0 | 51 | 0 | 26 | 1 | 56 | 0 | 57 | 50 |
| Glycan Biosynthesis and Metabolism | 62 | 3 | 57 | 0 | 67 | 1 | 34 | 0 | 68 | 1 | 70 | 61 |
| Metabolism of Cofactors and Vitamins | 79 | 0 | 69 | 0 | 89 | 1 | 38 | 1 | 85 | 0 | 82 | 61 |
| Biosynthesis of Polyketides and Terpenoids | 15 | 2 | 10 | 1 | 18 | 0 | 12 | 1 | 16 | 0 | 16 | 13 |
| Xenobiotics Biodegradation and Metabolism | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Transcription | 113 | 14 | 103 | 4 | 106 | 6 | 45 | 10 | 109 | 7 | 127 | 94 |
| Translation | 122 | 43 | 92 | 6 | 118 | 32 | 91 | 36 | 118 | 44 | 236 | 217 |
| Folding, Sorting and Degradation | 142 | 16 | 119 | 3 | 145 | 2 | 74 | 10 | 135 | 8 | 211 | 195 |
| Replication and Repair | 126 | 7 | 111 | 1 | 129 | 0 | 31 | 3 | 125 | 5 | 129 | 115 |
| Membrane Transport | 19 | 1 | 17 | 0 | 22 | 0 | 9 | 1 | 18 | 0 | 16 | 14 |
| Signal Transduction | 45 | 1 | 15 | 0 | 46 | 0 | 8 | 1 | 43 | 0 | 36 | 34 |
| Transport and Catabolism | 105 | 9 | 95 | 2 | 109 | 1 | 31 | 6 | 107 | 4 | 113 | 106 |
| Cell Growth and Death | 199 | 9 | 89 | 1 | 200 | 2 | 45 | 4 | 180 | 1 | 105 | 96 |
| *Total* | 1814 | 127 | 1489 | 25 | 1927 | 52 | 951 | 103 | 1843 | 79 | 2118 | 1883 |

Table 5.3: Intronic gene frequencies according to KEGG Pathways. *Genes* column denotes the total number of genes contained to the pathway. *Introns* column denotes the number of identified intronic genes contained in the pathway.

| **Cellular Pathway** | **PIC** | | **PPA** | | **SCE** | | **SPO** | | **YLI** | | **ZRO** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Genes* | *Introns* | *Genes* | *Introns* | *Genes* | *Introns* | *Genes* | *Introns* | *Genes* | *Introns* | *Genes* | *Introns* |
| Carbohydrate Metabolism | 331 | 37 | 173 | 23 | 333 | 2 | 225 | 62 | 264 | 75 | 191 | 2 |
| Energy Metabolism | 112 | 21 | 86 | 20 | 116 | 2 | 91 | 38 | 117 | 42 | 97 | 3 |
| Lipid Metabolism | 136 | 19 | 66 | 10 | 117 | 4 | 87 | 33 | 104 | 10 | 66 | 3 |
| Nucleotide Metabolism | 175 | 29 | 122 | 7 | 158 | 5 | 141 | 77 | 138 | 35 | 132 | 5 |
| Amino Acid Metabolism | 286 | 19 | 169 | 15 | 243 | 0 | 218 | 53 | 236 | 32 | 190 | 0 |
| Metabolism of Other Amino Acids | 88 | 10 | 46 | 1 | 59 | 0 | 51 | 11 | 57 | 6 | 43 | 0 |
| Glycan Biosynthesis and Metabolism | 70 | 18 | 45 | 4 | 79 | 2 | 52 | 23 | 56 | 8 | 59 | 1 |
| Metabolism of Cofactors and Vitamins | 104 | 18 | 64 | 4 | 100 | 0 | 87 | 25 | 83 | 6 | 81 | 0 |
| Biosynthesis of Polyketides and Terpenoids | 25 | 3 | 12 | 2 | 18 | 0 | 14 | 11 | 19 | 3 | 12 | 1 |
| Xenobiotics Biodegradation and Metabolism | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Transcription | 112 | 45 | 117 | 14 | 119 | 9 | 133 | 88 | 126 | 40 | 121 | 8 |
| Translation | 122 | 40 | 124 | 31 | 183 | 82 | 179 | 83 | 116 | 54 | 467 | 77 |
| Folding, Sorting and Degradation | 133 | 38 | 119 | 22 | 159 | 12 | 146 | 103 | 148 | 51 | 208 | 10 |
| Replication and Repair | 122 | 38 | 105 | 11 | 131 | 6 | 134 | 93 | 133 | 27 | 119 | 11 |
| Membrane Transport | 22 | 5 | 17 | 1 | 21 | 0 | 21 | 16 | 20 | 4 | 21 | 2 |
| Signal Transduction | 37 | 17 | 14 | 4 | 70 | 1 | 28 | 9 | 35 | 12 | 20 | 0 |
| Transport and Catabolism | 111 | 41 | 96 | 17 | 118 | 3 | 93 | 64 | 114 | 36 | 102 | 9 |
| Cell Growth and Death | 147 | 59 | 85 | 11 | 252 | 7 | 132 | 80 | 118 | 20 | 104 | 5 |
| *Total* | 2133 | 457 | 1460 | 197 | 2276 | 135 | 1832 | 869 | 1884 | 461 | 2033 | 137 |

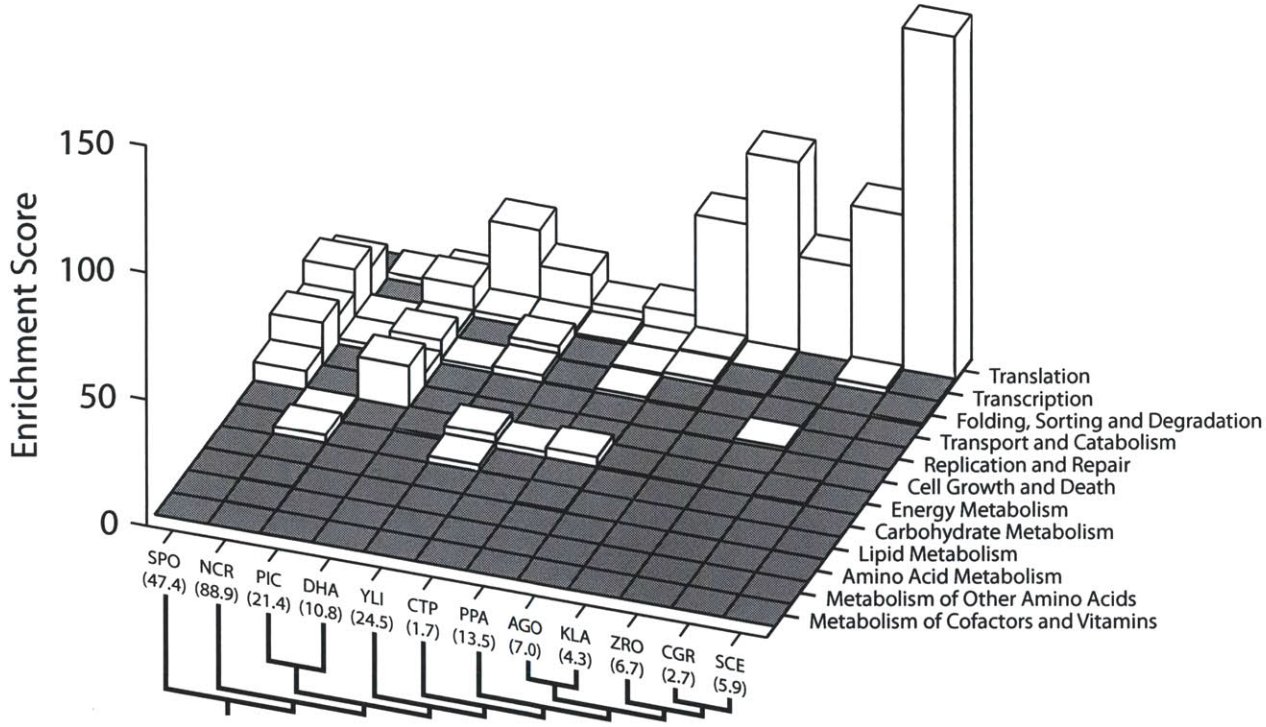


Figure 5.2.1: Intron-gene enrichment analysis of 12 ascomycetous yeast species and their phylogeny. Intronic gene distribution was tested for over-representation according to a hypergeometric distribution. Genes are categorized according to KEGG PATHWAY hierarchy. The enrichment score is calculated as the negative log of the positive tail p-value using Fisher Exact Test. Statistically significant enrichment scores (p-value $< 0.05$, white columns) are contrasted with non-enriched pathways (shaded columns). Below the horizontal axis is the phylogenic tree of analyzed yeast species (with percent of genes with introns below each name) generated using SUPERFAMILY (supfam.org); algorithm is based on neighbor-joining and maximum parsimony methods. Species abbreviations: SPO, *Schizosaccharomyces pombe*; NCR, *Neurospora crassa*; PIC, *Scheffersomyces stipitis*, formerly *Pichia stipitis*; DHA, *Debaryomyces hansenii*; YLI, *Yarrowia lipolytica*; CTP, *Candida tropicalis*; PPA, *Pichia pastoris*; AGO, *Ashbya gossypii*; KLA, Kluyveromyces lactis; ZRO, *Zygosaccharomyces rouxii*; CGR, *Candida glabrata*; SCE, *Saccharomyces cereivisiae.* Sources: Gene hierarchy assignments, KEGG (www.genome.jp/kegg/); intron lists: SCE, CGR, KLA, DHA, YLI, Genosplicing (genome.jouy.inra.fr/genosplicing/); SPO, Sanger Institute (www.sanger.ac.uk); All remaining intron lists and genomic data, NCBI (www.ncbi.nlm.nih.gov).

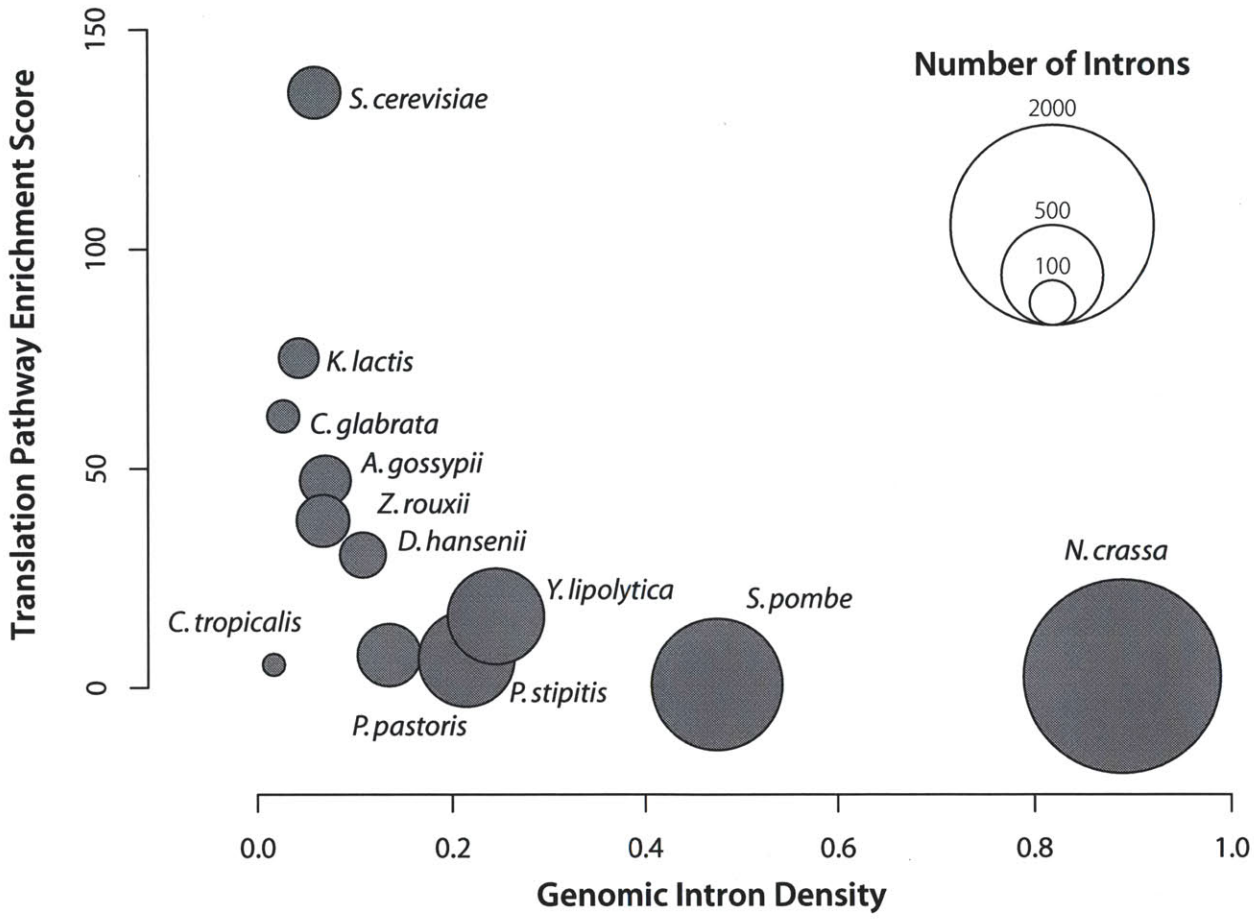


Figure 5.2.2: Comparison of intronic gene density in yeast genomes relative to translation pathway enrichment score. Intronic gene density calculated as the number of identified intronic genes divided by the number of pathway genes found in the KEGG database. Enrichment score is calculated as the negative log of the positive tail p-value using Fisher Exact Test. Area of each circle represents the number of intronic genes identified in the analysis.

## 5.3 Results & Discussion

The results obtained show two distinct patterns of intron enrichment: intron enrichment in a number of different pathways, and intron enrichment predominantly in translation-related genes. Combined with the phylogeny of the analyzed species, the observations suggest an evolutionary transition from diverse intron enrichment to specific enrichment of only translation-related genes. This trend parallels the transition from intron-rich to intron-poor genomes as observed in Figure 5.2.2, where we observe an inverse relationship between translation pathway enrichment score and overall intron density. Only *Candida tropicalis* exhibits an anomalous behavior: having both very low intronic gene density and low translation enrichment score. This could be attributed to either quality of intron annotation data available or as an environment-specific adaptation. *C. tropicalis* is known as a common human pathogen.

The enrichment of introns in translation pathway genes is also of particular note, as translation has been shown to be a prominent rate-limiting step in terms of growth rate (Scott et al., 2010). Combined with the observation that many intronic genes are highly expressed (Le Hir et al., 2003), this suggests that maximal expression of translation genes has been a key selected phenotype in order to optimize yeast towards higher growth rates. For example, in laboratory settings, *Saccharomyces cerevisiae* (intron-sparse, except for translation-associated pathways) grows roughly twice as fast as either *Schizosaccharomyces pombe* or *Neurospora crassa* (intron-rich, but not strongly enriched in translation-associated pathways) under optimal conditions (Jeffares et al., 2006).

One species-specific observation from the results is the intron enrichment in both carbohydrate and energy metabolism in the oleaginous yeast *Yarrowia lipolytica*. These categories encompass pathways for the catabolism of exogenous substrates and generation of ATP via oxidative phosphorylation. As an oleaginous yeast capable of utilizing of a wide-variety of carbon-rich substrates (Barth and Gaillardin, 1997), robust expression of pathways that generate energy from both stored and exogenous substrates was likely a selected phenotype giving rise to its current metabolic nature. A number of *Y. lipolytica* introns have been

validated as having expression enhancing properties, in particular the introns from fructose 1,6-bisphosphate aldolase (FBA1) and glyceraldehyde-3-phosphate dehydrogenase (TDH1) that were used to enhance expression for heterologous protein production (Hong et al., 2012). Our work showed that the intron from Translation Elongation Factor-1$\alpha$ (TEF) also exhibited expression enhancement. These genes were routinely identified as among the highest expression genes in the *Y. lipolytica* transcriptome, most likely due to these introns (Müller et al., 1998; Hong et al., 2012).

## 5.4 Conclusion

The use of bioinformatics allows us to elucidate the underlying relationships between intron function and genome evolution, uncovering connections that to date have been largely unexplored. Additionally, these analyses may also provide clues into more species-specific functional adaptations as introns enrich in specific pathways. Ultimately, if we recognize these intron distributions as broad genetic markers of yeast evolution, we discover a rich catalogue of the evolutionary prioritizations and optimizations these species undergo as they adapt to their ever-evolving environment.

# References


Barth, G., and Gaillardin, C. (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica. FEMS Microbiol. Rev. 19*, 219–237.

Callis, J., Fromm, M., and Walbot, V. (1987) Introns increase gene expression in cultured maize cells. *Genes & Development 1*, 1183–1200.

Choi, T., Huang, M., Gorman, C., and Jaenisch, R. (1991) A generic intron increases gene expression in transgenic mice. *Mol. Cell. Biol. 11*, 3070–3074.

Dujon, B. (2006) Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet. 22*, 375–387.

Fink, G. R. (1987) Pseudogenes in yeast? *Cell 49*, 5–6.

Hong, S.-P., Seip, J., Walters-Pollak, D., Rupert, R., Jackson, R., Xue, Z., and Zhu, Q. (2012) Engineering *Yarrowia lipolytica* to express secretory invertase with strong FBA1IN promoter. *Yeast 29*, 59–72.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res. 37*, 1–13.

Ivashchenko, A. T., Tauasarova, M. I., and Atambayeva, S. A. (2009) Exon-intron structure of genes in complete fungal genomes. *Molecular Biology 43*, 24–31.

Jeffares, D. C., Mourier, T., and Penny, D. (2006) The biology of intron gain and loss. *Trends Genet. 22*, 16–22.

Le Hir, H., Nott, A., and Moore, M. J. (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci. 28*, 215–220.

Lewin, B. *Genes IX*; Jones and Bartlett Publishers: Sudbury, MA, 2008.

Müller, S., Sandal, T., Kamp-Hansen, P., and Dalbøge, H. (1998) Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. *Yeast 14*, 1267–1283.

Neugebauer, K. M. (2002) On the importance of being co-transcriptional. *J. Cell Sci. 115*, 3865–3871.

Parenteau, J., Durand, M., Veronneau, S., Lacombe, A.-A., Morin, G., Guerin, V., Cecez, B., Gervais-Bird, J., Koh, C.-S., Brunelle, D., Wellinger, R. J., Chabot, B., and Abou Elela, S. (2008) Deletion of Many Yeast Introns Reveals a Minority of Genes that Require Splicing for Function. *Mol. Biol. Cell 19*, 1932–1941.

Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. (2010) Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science 330*, 1099–1102.

# Chapter 6

# Engineering the Push and Pull of Lipid Biosynthesis

## 6.1 Introduction

Liquid biofuels are a promising alternative to fossil fuels that can help ease concerns about climate change and smoothen supply uncertainties (Stephanopoulos, 2007). Biodiesel, jet oil and other oil-derived fuels in particular are necessary for aviation and heavy vehicle transport. Biodiesel is presently produced exclusively from vegetable oils, which is a costly and unsustainable path (Hill et al., 2006). For biodiesel, a transition from vegetable oil to microbial oil production for the oil feedstock presents numerous additional advantages: adaptability to diverse feedstocks, reduced land requirements, efficient process cycle turnover, and ease of scale-up (Beopoulos et al., 2011). Microbial platforms for bio-oil production are also more genetically tractable for further optimization.

Key to a cost-effective microbial technology for the conversion of carbohydrates to oils is high (carbohydrate to oil) conversion yields. Metabolic engineering has emerged as the enabling technology applied to this end (Tai and Stephanopoulos, 2012), building upon experience in successful pathway engineering of microbial biocatalysts for the synthesis of chemical, pharmaceutical and fuel products (Keasling, 2010). Prior efforts at engineering microbes with high lipid synthesis have focused on amplifying presumed rate-controlling steps in the fatty acid synthesis pathway (Courchesne et al., 2009). These efforts, however, have produced mixed results, as fatty acid synthesis tends to be tightly regulated in most organisms (Ohlrogge and Jaworski, 1997). Here we describe an approach that combines the amplification of upstream, metabolite-forming pathways with a similar increase in the flux of downstream, metabolite-consuming pathways. When balanced, this push-and-pull strategy can achieve large flux amplifications with minimal effects due to feedback inhibition.

The oleaginous yeast *Y. lipolytica* is an attractive candidate for microbial oil production, which has also been extensively used in a broad range of other industrial applications: citric acid production, protein production (i.e. proteases and lipases), and bioremediation (Scioli and Vollaro, 1997; Beckerich et al., 1998; Papanikolaou et al., 2002). With a fully sequenced genome and a growing body of tools, engineering of *Y. lipolytica* can be achieved with relative

ease (Barth and Gaillardin, 1997). It also has been found to be robust - able to grow on a variety of substrates - and has been used for lipid production on agro-industrial residues, industrial glycerol, and industrial fats (Papanikolaou and Aggelis, 2002, 2003; Papanikolaou et al., 2003). It has strong lipid accumulation capacity, commonly accumulating up to 36% of its dry cell weight (DCW) in lipids (Beopoulos et al., 2009).

The metabolic pathways for lipid synthesis in *Y. lipolytica* are beginning to be fully mapped out (see Figure 6.1.1). Transport of acetyl-CoA from the mitochondria to the cytosol is carried out by the ATP-citrate lyase (ACL)-mediated cleavage of citrate via the citrate shuttle yielding acetyl-CoA and oxaloacetate (OAA). Interestingly, this pathway for cytosolic acetyl-CoA generation has been found to be differentially present in oleaginous organisms (Boulton and Ratledge, 1981). Acetyl-CoA carboxylase (ACC) then catalyzes the first committed step towards lipid biosynthesis, converting cytosolic acetyl-CoA into malonyl-CoA, which is the primary precursor for fatty acid elongation. Completed fatty acyl-CoA chains are then transported to the endoplasmic reticulum (ER) or lipid body membranes for the final assembly of triacylglycerol (TAG) via the Kennedy pathway. Over 80% of the storage lipids produced in *Y. lipolytica* are in the form of TAG (Athenstaedt et al., 2006). Cytosolic OAA is converted to malate by malate dehydrogenase (MDH) and transported back into the mitochondria to complete the citrate shuttle cycle. Reducing equivalents in the form of NADPH are provided either by the pentose phosphate pathway (PPP) or by malic enzyme (MAE) in the pyruvate/OAA/malate transhydrogenase cycle. In *Y. lipolytica*, high PPP flux and ineffectual MAE overexpression suggest that the former is the primary source for NADPH (Blank et al., 2005; Beopoulos et al., 2011).

Intracellular lipid accumulation can occur via two methods: *de novo* lipid synthesis or *ex novo* incorporation of exogenous fatty acids and lipids. Lipid accumulation most commonly occurs when nutrient supply is exhausted in the presence of excess carbon, typically coinciding with the onset of stationary phase. In practice, the most commonly used limiting-nutrient is nitrogen, as it is easily controllable through media composition (Beopoulos et al.,

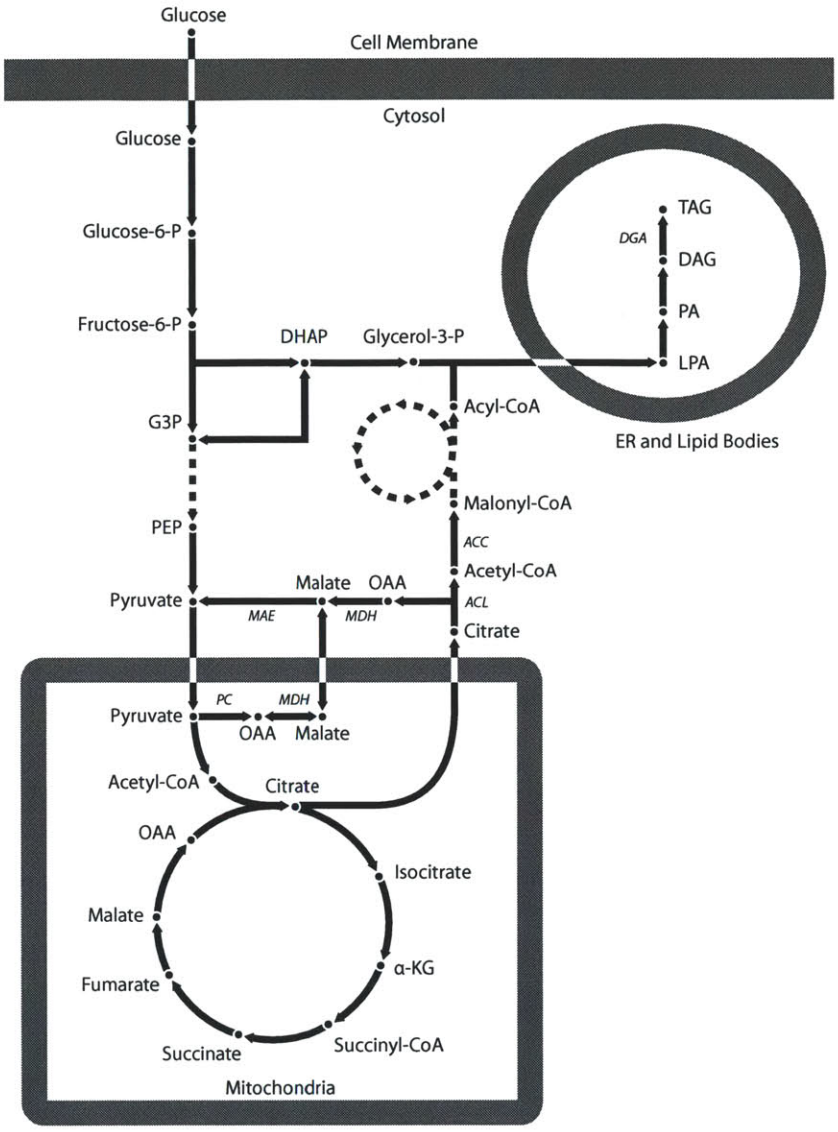


Figure 6.1.1: Overview of the principal metabolic pathways for lipid synthesis in *Y. lipolytica*. Glucose entering glycolysis enters the mitochondria as pyruvate for use in the TCA cycle; however, excess acetyl-coA is transported from the mitochondria to the cytosol via the citrate shuttle. Cytosolic acetyl-CoA is then converted into malonyl-CoA by acetyl-CoA carboxylase (ACC) as the first step of fatty acid synthesis. After fatty acid synthesis, triacylglycerol (TAG) synthesis follows the Kennedy pathway, which occurs in the endoplasmic reticulum (ER) and lipid bodies. Acyl-CoA is the precursor used for acylation to the glycerol-3-phosphate backbone to form lysophosphatidic acid (LPA), which is further acylated to form phosphatidic acid (PA). PA is then dephosphorylated to form diacylglycerol (DAG) and then a final acylation occurs by diacylglycerol acyltransferase (DGA) to produce TAG. OAA oxaloacetate, $\alpha$-KG alpha-ketoglutarate, PEP phosphoenolpyruvate, G3P Glyceraldehyde 3-phosphate, DHAP dihydroxyacetone phosphate, ACL ATP citrate lyase, PC pyruvate carboxylase, MDH malate dehydrogenase, MAE malic enzyme.

2009). Ultimately, lipid synthesis pathways are highly regulated in order for the organism to balance cell growth with energy storage. For example, ACC alone is regulated at multiple levels and by multiple factors (Ohlrogge and Jaworski, 1997).

Engineering lipid biosynthesis pathways in *Y. lipolytica* is still relatively unexplored, though initial attempts have shown promise. By eliminating peroxisomal oxidation pathways and engineering glycerol metabolism, *Y. lipolytica* was able to achieve 40%-70% lipids through *ex novo* lipid accumulation (Beopoulos et al., 2008; Dulermo and Nicaud, 2011). Co-expression of $\Delta$6- and $\Delta$12-desaturase genes allowed for significant production of $\gamma$-linolenic acid (GLA) (Chuang et al., 2010). Strategies continue to develop for effective engineering of the lipid production pathways to maximize output. By combining the tools for genetic engineering of *Y. lipolytica* with tested strategies developed in the practice of metabolic engineering, we can achieve significant increases of lipid production in this oleaginous yeast host. Here we explore the effects of ACC1 and DGA1 overexpression on lipid accumulation via *de novo* lipid biosynthesis. The coupling of ACC1 and DGA overexpression allows effective flux diversion towards lipid synthesis and creation of a driving force by sequestering product formation in lipid bodies.

## 6.2 Materials and Methods

### 6.2.1 Yeast strains, growth, and culture conditions

The *Y. lipolytica* strains used in this study were derived from the wild-type *Y. lipolytica* W29 strain (ATCC20460). The auxotrophic Po1g (Leu-) used in all transformations was obtained from Yeastern Biotech Company (Taipei, Taiwan). All strains used in this study are listed in Table 6.1.

Media and growth conditions for *Escherichia coli* have been previously described by Sambrook et al. (2001), and those for *Y. lipolytica* have been described by Barth and Gaillardin (1997). Rich medium (YPD) was prepared with 20 g/L Bacto peptone (Difco

Table 6.1: Strains and plasmids used in this study

| Strains (host strain) | Genotype or plasmid | Source |
|---|---|---|
| *E. coli* | | |
| DH5$\alpha$ | fhuA2 $\Delta$(argF-lacZ)U169 phoA glnV44 $\Phi$80 $\Delta$(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 | Invitrogen |
| pINA1269 | JMP62-LEU | Yeastern |
| pMT010 | pINA1269 php4d::TEF | Chapter 4 |
| pMT015 | pINA1269 php4d::TEFin | Chapter 4 |
| pMT013 | YTEF-ACC1 | This work |
| pMT040 | php4d-ACC1 | This work |
| pMT053 | YTEFin-DGA | This work |
| pMT065 | php4d-ACC1 + YTEFin-DGA | This work |
| *Y. lipolytica* | | |
| Po1g | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 | Yeastern |
| MTYL037 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 YTEFin-LacZ | Chapter 4 |
| MTYL040 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 hp4d-ACC1-LEU2 | This work |
| MTYL053 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEFin-DGA1-LEU2 | This work |
| MTYL065 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 hp4d-ACC1 + TEFin-DGA1-LEU2 | This work |

Table 6.2: Primers used in this study. Relevant restriction sites are in bold.

| Primer | Description | Sequence |
|---|---|---|
| *PCR* | | |
| MT080 | TEF-ACC | *GACT**ACGCGT**CACAATGCGACTGCAATTGA* |
| MT081 | TEF-ACC | *TAGC**ATGCAT**TCACAACCCCTTGAGCAGCT* |
| MT222 | Hp4d-ACC1 | *AATGCGACTGCAATTGAGGACACTAA* |
| MT137 | Hp4d-ACC1 | *CGTTGAATC**GAATTC**GGACACGGGCATCTCAC* |
| MT271 | DGA | *TAACCGCAGACTATCGACTCACAATACTACAAGTCGCG* |
| MT272 | DGA | *CTAGGT**ATGCAT**TTACTCAATCATTCGGAACTCTGGG* |
| MT220 | Cassette | *CCCGGCAACA**ATTAAT**AGACTGGAT* |
| MT265 | Cassette | *TTCGGACACGGGCATCTCAC* |
| *RT-PCR* | | |
| MTR001 | Actin | *TCCAGGCCGTCCTCTCCC* |
| MTR002 | Actin | *GGCCAGCCATATCGAGTCGCA* |
| MTR031 | DGA | *AACGGAGGAGTGGTCAAGCGA* |
| MTR032 | DGA | *TTATGGGGAAGTAGCGGCCAA* |
| MTR041 | ACC | *GCTTCTCACGGAGGTCATACAGT* |
| MTR042 | ACC | *CTGCGGCAATACCGTTGTTAG* |

Laboratories, Detroit, MI), 10 g/L yeast extract (Difco), 20 g/L glucose (Sigma-Aldrich, St. Louis, MO). YNB medium was made with 1.7 g/L yeast nitrogen base (without amino acids) (Difco), 0.69 g/L CSM-Leu (MP Biomedicals, Solon, OH), and 20 g/L glucose. Selective YNB plates contained 1.7 g/L yeast nitrogen base (without amino acids), 0.69 g/L CSM-Leu, 20 g/L glucose, and 15 g/L Bacto agar (Difco).

Shake flask experiments were carried out using the following medium: 1.7 g/L yeast nitrogen base (without amino acids), 1.5 g/L yeast extract, and 50 g/L glucose. From frozen stocks, precultures were inoculated into YNB medium (5 mL in Falcon tube, 200 rpm, 28°C, 24 hr). Overnight cultures were inoculated into 50 mL of media in 250 mL Erlenmeyer shake flask to an optical density ($A_{600}$) of 0.05 and allowed to incubate for 100 hours (200 rpm, 28°C), after which biomass, sugar content, and lipid content were taken and analyzed.

Bioreactor scale fermentation was carried out in a 2-liter baffled stirred-tank bioreactor. The medium used contained 1.5 g/L yeast nitrogen base (without amino acids and ammonium sulfate), 2 g/L ammonium sulfate, 1 g/L yeast extract, and 90 g/L glucose. From a selective plate, an initial preculture was inoculated into YPD medium (40 mL in 250 mL Erlenmeyer flask, 200 rpm, 28°C, 24 hr). Exponentially growing cells from the overnight preculture were transferred into the bioreactor to an optical density ($A_{600}$) of 0.1 in the 2-L reactor (2.5 vvm aeration, pH 6.8, 28°C, 250 rpm agitation). Time point samples were stored at -20°C for subsequent lipid analysis. Sugar organic acid content was determined by HPLC. Biomass was determined by determined gravimetrically from samples dried at 60°C for two nights.

### 6.2.2 Plasmid construction

Standard molecular genetic techniques were used throughout this study (Sambrook and Russell, 2001). Restriction enzymes and Phusion High-Fidelity DNA polymerase used in cloning were obtained from New England Biolabs (Ipswich, MA). Genomic DNA from yeast transformants was prepared using Yeastar Genomic DNA kit (Zymo Research, Irvine, CA).

All constructed plasmids were verified by sequencing. PCR products and DNA fragments were purified with PCR Purification Kit or QIAEX II kit (Qiagen, Valencia, CA). Plasmids used are described in Table 6.1. Primers used are described in Table 6.2.

Plasmid pMT010 was constructed by amplifying the translation elongation factor-1$\alpha$ (TEF) promoter region (Accession number: AF054508) from *Y. lipolytica* Po1g genomic DNA using primers MT078 and MT079. The amplicon was inserted between SalI and KpnI sites of the starting vector, pINA1269, also known as pYLEX1, obtained from Yeastern Biotech Company (Taipei, Taiwan). Also included in the reverse primer MT079 were MluI and NsiI sites to add restriction sites to the multi-cloning site.

Plasmid pMT015 was constructed by amplifying from *Y. lipolytica* Po1g genomic DNA the TEF promoter and the 5' coding region containing the ATG start codon and 113 bp of the endogenous intron (Accession number: CR382129). Primers MT118 and MT122 were used for this amplification and inserted between SalI and MluI sites of pMT010. For cloning purposes, some of the intron was omitted so that the SnaBI restriction site could be incorporated. Cloning a gene into this plasmid thus requires the omission of the gene's ATG start codon, addition of TAACCGCAG to the beginning of the 5' primer, and blunt-end ligation at the 5' end.

Plasmid pMT025 was constructed by amplifying the LacZ gene, encoding $\beta$-galactosidase, from *E. coli* and inserting it into the PmlI and BamHI sites of starting vector pINA1269 using primers MT170 and MT171. Plasmid pMT038 was constructed by amplifying the LacZ gene and inserting it into the MluI and NsiI sites of pMT010 using primers MT168 and MT169. Since LacZ contains multiple MluI sites, AscI was used as the 5' restriction site on MT168 that has a matching overhang. Plasmid pMT037 was constructed by amplifying LacZ gene and inserting it into the SnaBI and NsiI sites of pMT015. Primers MT172 and MT169 were used, where forward primer MT127 omits the ATG start codon of LacZ and instead begins with the sequence TAACCGCAG that completes the intron sequence of pMT015.

Plasmid pMT013 was constructed by amplifying the ACC1 gene from *Y. lipolytica* Po1g genomic DNA (Accession Number: XM_501721) and inserting it into the MluI and NsiI sites of pMT010 using primers MT080 and MT081. Plasmid pMT040 was constructed by amplifying the ACC1 gene and its terminator from pMT013 using primers MT222 and MT137 and inserting this into starting vector pINA1269 digested with PmlI and ClaI.

Plasmid pMT053 was constructed by amplifying the DGA1 gene from *Y. lipolytica* Po1g genomic DNA (Accession Number: XM_504700) using primers MT271 and MT272. The amplified gene was digested with NsiI and was inserted into PMT015 in the same manner as in the construction of pMT037.

To produce a single plasmid that could express both ACC1 and DGA1, a promoter-gene-terminator cassette was amplified from pMT053 using primers MT220 and MT265. This was then digested with DpnI and AseI and inserted into pMT040 that was digested with NruI and AseI resulting in tandem gene construct pMT065. The AseI restriction site was selected to facilitate selection, as it resides within the Ampicillin resistance marker. Because NruI is a blunt end restriction site, insertion of the amplicon does not increase the total number of NruI sites to facilitate progressive insertions.

Plasmids were linearized with either NotI or SacII and chromosomally integrated into Po1g according to the one-step lithium acetate transformation method described by Chen et al. (1997). Transformants were plated on selective media and verified by PCR of prepared genomic DNA. Verified transformants were then stored as frozen stocks at -80°C and on selective YNB plates at 4°C.

### 6.2.3 RNA isolation and transcript quantification

Shake flask cultures grown for 42 hrs were collected and centrifuged for 5 min at 10,000g. Each pellet was resuspended in 1.0 ml of Trizol reagent (Invitrogen) and 100 $\mu$L of acid-washed glass beads were added (Sigma-Aldrich). Tubes were vortexed for 15 min at 4°C for cell lysis to occur. The tubes were then centrifuged for 10 min at 12,000g at 4°C and the

supernatant was collected in a fresh 2-mL tube. 200 $\mu$L chloroform was then added and tubes were shaken by hand for 10 seconds. The tubes were again centrifuged for 10 min at 12,000g at 4°C. 400 $\mu$L of the upper aqueous phase was transferred to a new tube, and an equal volume of phenol-chloroform-isoamyl alcohol (pH 4.7) (Ambion, Austin, TX) was added. Tubes were again shaken by hand for 10 seconds and centrifuged for 10 min at 12,000g at 4°C. 250 $\mu$L of the upper phase was transferred to a new tube with an equal volume of cold ethanol and 1/10th volume sodium acetate (pH 5.2). Tubes were chilled at -20°C for thirty minutes to promote precipitation. Tubes were then centrifuged for 5 min at 12,000g, washed twice with 70% ethanol, dried in a 60°C oven and finally resuspended in RNAse free water. RNA quantity was analyzed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and samples were stored in -80°C freezer. qRT-PCR analyses were carried out using iScript One-step RT-PCR Kit with SYBR Green (Bio-Rad, Hercules, CA) using the Bio-Rad iCycler iQ Real-Time PCR Detection System. Fluorescence results were analyzed using Real-time PCR Miner and relative quantification and statistical analysis was determined with REST 2009 (Qiagen) using actin as the reference gene and MTYL038 as the reference strain (Zhao and Fernald, 2005). Samples were analyzed in quadruplicate.

### 6.2.4 Lipid extraction and quantification

Total lipids were extracted using the procedure by Folch et al (Folch et al., 1957). A measured quantity of cell biomass (roughly 1 mg) was suspended in 1 mL of chloroform:methanol (2:1) solution and vortexed for 1 hour. After centrifugation, 500 $\mu$L was transferred to 125 $\mu$L saline solution. The upper aqueous layer was removed and the bottom layer was evaporated and resuspend in 100 $\mu$L hexane. Samples were then stored at -20°C until transesterification.

Transesterification of total lipid extracts was performed by adding 1 mL 2% (wt/vol) sulfuric acid in methanol to each sample. Samples were then incubated at 60°C for 2 hours. After that the samples were partially evaporated, and the fatty acid methyl esters (FAME)

Table 6.3: Total fatty acid content, yield and distribution for *Y. lipolytica* strains. Total fatty acid content is given as means ± S.D. (n = 3) for a 50 mL culture after 100 hrs (C/N molar ratio of 20). Fatty acid profiles are given as percent of fatty acid of total fatty acids, with error less than 2.5%.

| | | | | Fatty Acid (Fractional Abundance %) | | | | |
|---|---|---|---|---|---|---|---|---|
| | Biomass (g DCW) | Lipid Content (%) | Oil Yield (g/100 g glucose) | C16 | C16:1 | C18:0 | C18:1 | C18:2 |
| Control (MTYL037) | 10.49 | 8.77 ± 0.37 | 2.28 | 20.5 | 3.9 | 24 | 43.0 | 8.7 |
| ACC1 (MTYL040) | 6.04 | 17.9 ± 1.13 | 6.25 | 20.3 | 3.5 | 23.4 | 41.7 | 11.0 |
| DGA1 (MTYL053) | 7.53 | 33.8 ± 0.55 | 9.36 | 18.3 | 2.6 | 34.1 | 38.6 | 6.4 |
| ACC1+DGA1 (MTYL065) | 8.6 | 41.4 ± 1.90 | 11.4 | 16.0 | 2.3 | 32.8 | 44.9 | 4.0 |

were extracted by adding 1 mL hexane and vortexing for 10 min. 800 $\mu$L of this hexane was then transferred into glass vials for GC analysis.

GC analysis of FAMEs was performed with a Bruker 450-GC instrument equipped with a flame-ionization detector and a capillary column HP-INNOWAX (30 m × 0.25 mm). The GC oven conditions were as follows: 150°C (1 min), a 10 min ramp to 230°C, hold at 230°C for 2 min. The split ratio was 10:1. Fatty acids were identified and quantified by comparison with commercial FAME standards normalized to methyl tridecanoate (C13:0). Total lipid content was calculated as the sum of total fatty acid contents for five FAMEs: methyl palmitate (C16:0), methyl palmitoleate (C16:1), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2) (Sigma-Aldrich). The addition of tridecanoic acid to the chloroform-methanol extraction fluid was used as the internal standard, which was carried through the entire analysis procedure and transesterified into its methyl ester.

## 6.3 Results & Discussion

### 6.3.1 Overexpression of ACC1 and DGA1 leads to significant increases in lipid accumulation

The use of the TEF promoter along with its expression-enhancing intron provides an excellent platform for high gene expression in *Y. lipolytica* (See Chapter 4 on page 59). We therefore

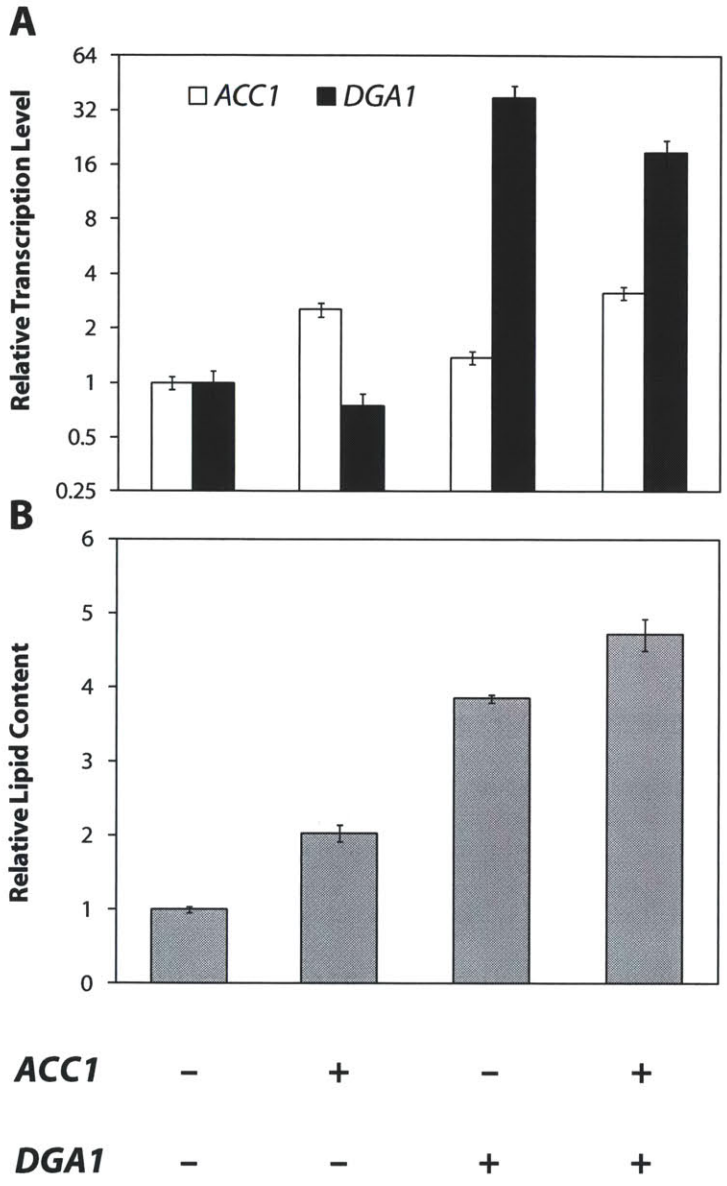


Figure 6.3.1: Combinatorial analysis of strains overexpressing ACC1 and/or DGA1. (A) Relative quantification of RNA transcripts using RT-PCR. Actin was used as the reference gene. Samples performed in quadruplicate, standard error estimated using REST 2009 software. (B) Relative lipid content of strains through total fatty acid analysis after 100 hrs of culture (C/N Ratio molar ratio of 20). Lipid samples performed in triplicate. MYTL038 was used as the reference strain.

used this for the overexpression of DGA1 (pMT053), which has been shown to be important in lipid accumulation in both *Y. lipolytica* and *S. cerevisiae* (Kamisaka et al., 2007; Dulermo and Nicaud, 2011). ACC1, already having two endogenous promoter-proximal introns, was not cloned with the TEFin promoter. Instead it was cloned with TEF and hp4d promoters (pMT013 and pMT040, respectively). Growth rates and lipid production were relatively similar between the two (data not shown), so php4d-ACC1 (pMT040) was used for lipid experiments and for the tandem gene construction of ACC1+DGA1 (pMT065). The hp4d promoter also was selected to minimize the possibility of homologous recombination of the two parallel gene cassettes in the ACC1+DGA1 construct. Simultaneous coexpression of two genes in *Y. lipolytica* using tandem gene construction has been successfully performed elsewhere (Chuang et al., 2010).

Plasmid integration was verified by PCR of prepared genomic DNA of the transformed strains. Overexpression was also measured by RT-PCR of total RNA (Figure 6.3.1A). It was found that the transformants indeed exhibit increased expression of ACC1 and DGA1 in the respective strains, with DGA1 being much more highly expressed than ACC1, up to 40-fold increase in expression over the control. ACC1 was only overexpressed 3-fold in the transformants, possibly due to the relatively large transcript size of the gene (7,000 bp). Varying the promoter did not demonstrate any difference in expression for ACC1 or DGA1. Slight up-regulation of ACC1 in the DGA1 transformant was also observed, suggesting a possible regulatory relationship between the two genes.

The effect of ACC1 and DGA1 overexpression on lipid production was first assessed in shake flask experiments (Figure 6.3.1B and Table 6.3). The control LacZ strain only produced 8.77% (g/g DCW) lipids, which is similar to wild-type performance in shake flasks with glucose as sole substrate (Papanikolaou et al., 2006). ACC1 and DGA1 transformants both outperformed the control, accumulating 17.9% and 33.8% lipid content, respectively. DGA1 in particular exhibited almost twice as much lipid accumulation as ACC1, almost 4-fold over the control. The biomass generated from the control was significantly higher than

the other strains, suggesting that the expression of ACC1 and DGA1 impairs somewhat the growth of *Y. lipolytica* while diverting cellular resources towards oil synthesis rather than growth. Overall oil yields were relatively low in comparison to a theoretical maximum yield of 0.32 g/g (Ratledge, 1988). There were also slight shifts in the fatty acid profile, with ACC1 producing significantly more linoleic acid and DGA1 maintaining a higher proportion of stearic acid. The proportion of oleic acid stayed relatively even across all transformants.

Improving upon both single gene transformants, ACC1+DGA1 was able to achieve 41.4% lipid content, a 4.7-fold improvement over the control. The biomass production was also improved over the single transformants, but still less than the control. Oil yield improved proportionally, to 0.114 g/g, or 35% of theoretical yield.

Compared to other similar work, the lipid content of the control strain is considered relatively low. However, the media composition in the shake flask experiments had a C/N molar ratio of only 20, and lipid accumulation was derived solely from *de novo* synthesis. Optimal C/N molar ratios for lipid accumulation typically range from 60-100 (Beopoulos et al., 2009), suggesting that higher lipid contents can be achieved under optimized conditions.

In eukaryotic organisms, overexpression of ACC has been met with only limited improvement of lipid production. Heterologous expression of ACC1 from the oleaginous fungus *Mucor rouxii* in the non-oleaginous yeast *Hansenula polymorpha* was able to achieve only a 40% increase in total fatty acid content, from 3.8% to 5.3% (Ruenwai et al., 2009). In plants, overexpression of *Arabidopsis* ACC1 led to dramatic increases in enzyme activity, but to no more than 30% increase in final lipid content (Roesler et al., 1997; Klaus et al., 2004). It is suspected that improvements in total lipid accumulation have been limited in eukaryotes by the strong metabolic and regulatory control maintained over this enzyme, possibly by free fatty acids. ACC expression and activity is influenced by numerous transcription factors, protein kinases, and metabolites (Brownsey et al., 2006). For example, in *Candida* (*Yarrowia*) *lipolytica*, the accumulation of acyl-CoA in acetyl-CoA synthetase mutants led to an 8-fold decrease in ACC activity (Kamiryo et al., 1979). Nonetheless, *Y. lipolytica*

might represent a regulatory exception in eukaryotic organisms, lending much to its oleaginous nature, as here we achieve a 2-fold increase in lipid content through a commensurate overexpression of endogenous ACC1.

The role of DGA has only recently been emphasized to be important for growth and lipid synthesis. In *Y. lipolytica*, three acyltransferases perform the final step of converting diacylglycerol (DAG) into triacylglycerol (DGA1/DGA2/PDAT). A triple knockout of these acyltransferases results in significant growth defect in both lag phase and growth rate, suggesting connections between oil synthesis and normal growth (Zhang et al., 2011). Interestingly, transcriptomic analysis of *Y. lipolytica* identifies only DGA2 as differentially expressed in lipid accumulation phases, suggesting activation of lipid accumulation may only partially be controlled at the transcription level (Morin et al., 2011). Nonetheless, DGA1p has been identified to predominantly localize to the membrane surface of lipid bodies and is thought to act in concert with triglyceride lipase (TGL3) to balance TAG flux in and out of lipid bodies (Athenstaedt et al., 2006). One thus expects the storage of TAGs to rest heavily on the relative activity (and abundance) of DGA1p with respect to its TGL3p counterpart. It has also been hypothesized that DGA diverts flux away from phospholipid synthesis thus creating a driving force for lipid synthesis, as an increased flux is required to produce the still necessary phospholipids (Courchesne et al., 2009). Whether this dynamic and resulting phenotype occurs with overexpression of the other two acyltransferases is unclear, as localization and protein-level interactions may play a role in their regulation. DGA1 overexpression in other organisms has also led to significant improvements: in an oleaginous $\Delta$snf2 *S. cerevisiae* mutant, DGA1 overexpression led to accumulation of up to 27% lipid content, a 2.3-fold increase (Kamisaka et al., 2007); in *Arabidopsis*, DGAT overexpression led to a 20-fold increase in lipid content in the leaves, and two-fold overall (Andrianov et al., 2010).

The enhanced lipid accumulation observed in the strains co-expressing ACC1 and DGA1 is presumably due to a better balance between the fatty acid and TAG synthesis path-

ways. Acyl-CoA intermediates function as both product and feedback inhibitors in the fatty acid (upstream) pathway and primary precursors in the TAG (downstream) pathway. Up-regulation of the upstream pathway increases the throughput of fatty acid synthesis and, for ACC in particular, diverts flux away from any pathways that would compete for cytosolic acetyl-CoA. Up-regulation of the downstream pathway creates a driving force by depleting acyl-CoA intermediates and increasing the rate of storage of TAG in lipid bodies. However, the two pathways must be well balanced because if modulated individually, they can lead to imbalances that can produce adverse effects on cell metabolism and growth.

Coexpressing ACC1 and DGA1 allows for increased throughput simultaneously in both upstream and downstream pathways while minimizing intermediate metabolite accumulation. This leads to increased lipid production as it combines high flux through lipid synthesis from ACC with the driving force provided by the sequestration of TAG into lipid bodies by DGA. The result is a synergistic increase in lipid accumulation, almost 5-fold greater than the control. Indeed, coupling precursor overproduction and driving forces with a metabolic sink to enable a push-and-pull dynamic has become a very powerful strategy in recent efforts of metabolic engineering, particularly for biofuels (Huo et al., 2011; Liu et al., 2011; Shen et al., 2011).

### 6.3.2 Fermentation performance of the ACC1 + DGA1 transformant

To further characterize the ACC1+DGA1 transformant (MTYL065) and explore its lipid accumulation characteristics, large-scale fermentation was conducted using a 2-L stirred-tank bioreactor. Glucose concentration was increased and ammonium sulfate concentration reduced to achieve a C/N molar ratio of 100.

The strain MTYL037 was used as a control (Figure 6.3.2A), consuming the 90 g/L of glucose within 120 hrs of fermentation. The culture quickly grew to a biomass concentration of 20 g/L within 48 hrs, but biomass accumulation increased very little afterwards. Over the

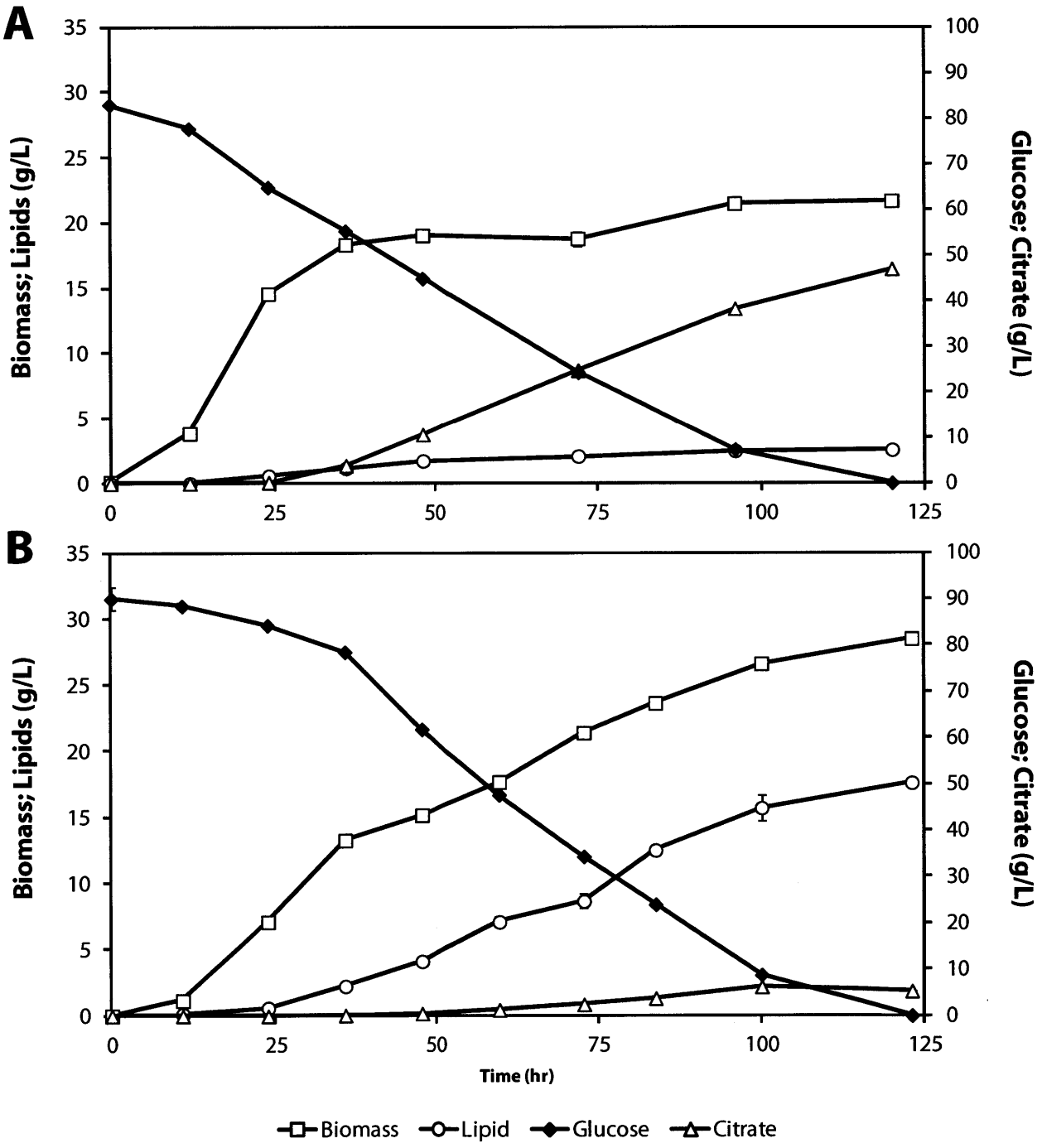


Figure 6.3.2: Batch bioreactor fermentation of control (A) LacZ transformant (MTYL037) compared to (B) ACC1+DGA1 transformant (MTYL065). C/N molar ratio was 100. All sampling was performed in triplicate.

course of the fermentation, only 2.5 g/L of the 21.6 g/L of biomass was in the form of lipids, resulting in 11.7% total lipid content. Instead, a majority of the carbon was converted into citrate, generating 47 g/L by the end of the fermentation and beginning near the onset of the stationary phase.

For strain MTYL065, glucose was fully consumed over the course of the 120 hour fermentation, with final biomass reaching 28.5 g/L (Figure 6.3.2B). Final lipid content was 61.7% of DCW, a 50% increase in lipid accumulation compared to the shake flask experiment, and a 5-fold increase over the control bioreactor, similar to the ratio observed in the shakeflasks. Furthermore, citrate production was lower, producing only about 5 g/L despite the same C/N ratio. These results compare favorably with other sugar fermentations (Aggelis and Komaitis, 1999; Karatay and Dönmez, 2010; Cui et al., 2011; Tsigie et al., 2011), as well as values found in *ex novo* lipid accumulation schemes (Papanikolaou et al., 2002; Dulermo and Nicaud, 2011). Overall yield and productivity was 0.195 g/g and 0.143 g/L/hr, respectively; however, during maximum lipid production observed between 70-100 hours, a yield of 0.270 g/g and a productivity of 0.178 g/L/hr were reached. Almost all biomass produced during this phase can be accounted for by the increase in lipid content. The overall and maximum yields achieved are 60.9% and 84.3% of the theoretical yield for TAG synthesis.

While the fatty acid profile is similar between MTYL037 and MTYL065 at the shake flask level, the fatty acid profile changed dramatically during the scale-up (Figure 6.3.3). Relative depletion of stearic acid and accumulation of palmitic acid was observed, with the predominant oleic acid ultimately comprising 49.3% of total fatty acids. The ratio between oleic and stearic acids steadily increased throughout the fermentation (data not shown), finally ending in a ratio of 4.6. This is a dramatic change from the ratio of 1.3 seen in shake flask experiments.

High relative oleic acid concentrations (up to 58.5%) and oleate:stearate ratios have also been observed in other 2-L fermentations (Zhao et al., 2010). This is similar to profiles of other oleaginous yeasts that accumulate more than 50% lipid content (Beopoulos et al.,

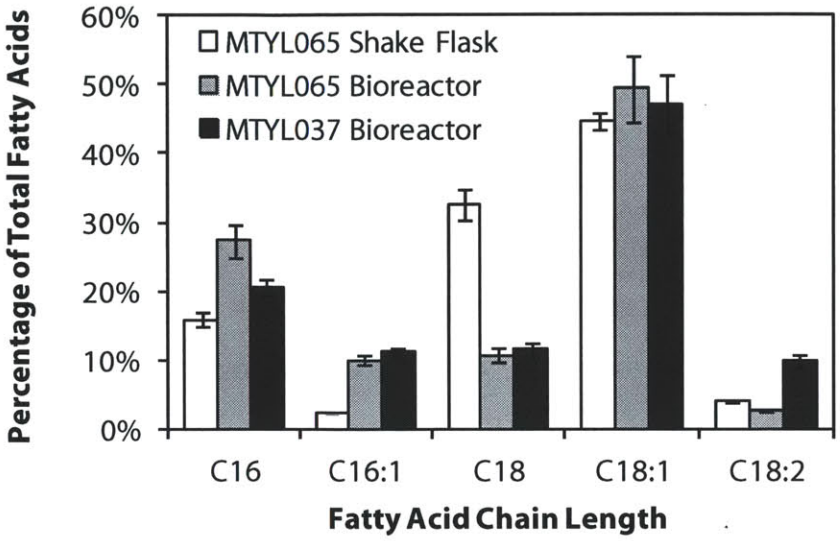


Figure 6.3.3: Fatty acid distribution comparisons for MTYL065 grown in shake flask, 2-L bioreactor fermentation and control strain MTYL037 in 2-L bioreactor fermentation. Samples performed in triplicate. C16 palmitate; C16:1 palmitoleate; C18 stearate; C18:1 oleate; C18:2 linoleate.

2009). In conditions of rapid lipid production, oleic acid might be more rapidly stored and easier to accumulate, as DGA1p is known to have varying specificities for different acyl-CoA; in *S. cerevisiae*, C18:1 and C16:0 are the more preferred substrates, having twice the activity of C18:0 (Oelkers et al., 2002). Furthermore, the high oleic acid concentration might also be a response to the higher aeration rate or reactor stresses of the bioreactor and not normally found in shake flask fermentations.

The relative reduction in citrate accumulation may also be a product of the metabolic engineering, as C/N ratios of 100 are normally known to induce citrate accumulation over lipid production. This is observed in our control bioreactor fermentation as well as in literature (Beopoulos et al., 2009). The driving force generated from ACC1+DGA1 overexpression seems to minimize accumulation of citrate. This both improves flux towards lipid production as well as provides greater flexibility in the use of C/N ratio to control lipid synthesis.

The high lipid content, yield, and productivity seen in the ACC1+DGA1 strain demonstrate the innate capacity of *Y. lipolytica* to accommodate high flux through the lipid syn-

thesis pathway. With further modifications and process optimization, *Y. lipolytica* can yield promising breakthroughs in the robust, efficient *de novo* synthesis of lipids.

## 6.4 Conclusions

Lipid biosynthesis is a tightly regulated metabolic pathway. For industrially-relevant applications of microbial lipid production, effective engineering of biosynthetic pathways is necessary to maximize yields and productivity. The use of the oleaginous yeast *Y. lipolytica* benefits from its high capacity for lipid accumulation and well-developed tools for engineering the lipid metabolic pathway. Here we show that the co-overexpression of two important genes in the lipid synthesis pathway, ACC1 and DGA1, provides an enhanced driving force towards the production of lipids, under both moderate and extreme C/N ratios. As the two enzymes carry out the first and last steps of lipid synthesis, the simultaneous push-and-pull of carbon flux towards TAG allows for enhanced production with minimal feedback inhibition. The resulting ACC1+DGA1 strain was able to accumulate up to 62% of its DCW as lipids through *de novo* synthesis at an overall volumetric productivity of 0.143 g/L/hr.

The concepts of (a) strong overexpression of pathway genes, (b) balance of upstream and downstream pathways, (c) diversion of flux towards desired pathways, and (d) driving forces towards the final product, are prominent strategies in the practice of metabolic engineering, where metabolic networks are engineered and optimized for the generation of desirable products. Implementation of these concepts with respect to lipid accumulation may readily extend to a number of biological platforms, including microalgae. These strategies will be foundational in enabling the technologies of robust, efficient, commodity-scale production of biologically-derived chemicals and fuels. Their application to lipid biosynthesis opens the path for microbial oil overproduction and cost-effective biofuel manufacturing.

# References

Aggelis, G., and Komaitis, M. (1999) Enhancement of single cell oil production by *Yarrowia lipolytica* growing in the presence of Teucrium polium L. aqueous extract. *Biotechnol. Lett. 21*, 747–749.

Andrianov, V., Borisjuk, N., Pogrebnyak, N., Brinker, A., Dixon, J., Spitsin, S., Flynn, J., Matyszczuk, P., Andryszak, K., and Laurelli, M. (2010) Tobacco as a production platform for biofuel: overexpression of *Arabidopsis* DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. *Plant Biotechnology Journal 8*, 277–287.

Athenstaedt, K., Jolivet, P., Boulard, C., Zivy, M., Negroni, L., Nicaud, J. M., and Chardot, T. (2006) Lipid particle composition of the yeast *Yarrowia lipolytica* depends on the carbon source. *Proteomics 6*, 1450–1459.

Barth, G., and Gaillardin, C. (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica. FEMS Microbiol. Rev. 19*, 219–237.

Beckerich, J. M., Boisramé-Baudevin, A., and Gaillardin, C. (1998) *Yarrowia lipolytica*: a model organism for protein secretion studies. *International Microbiology 1*, 123.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J. L., Molina-Jouve, C., and Nicaud, J. M. (2009) *Yarrowia lipolytica* as a model for bio-oil production. *Progress in Lipid Research 48*, 375–387.

Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M. T., Hapala, I., Papanikolaou, S., Chardot, T., and Nicaud, J. M. (2008) Control of lipid accumulation in the yeast *Yarrowia lipolytica. Appl. Environ. Microbiol. 74*, 7779.

Beopoulos, A., Nicaud, J.-M., and Gaillardin, C. (2011) An overview of lipid metabolism in yeasts and its impact on biotechnological processes. *Appl. Microbiol. Biotechnol. 90*, 1193–1206.

Blank, L. M., Lehmbeck, F., and Sauer, U. (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res. 5*, 545–558.

Boulton, C. A., and Ratledge, C. (1981) Correlation of Lipid Accumulation in Yeasts with Possession of ATP: Citrate Lyase. *Journal of General Microbiology 127*, 169–176.

Brownsey, R. W., Boone, A. N., Elliott, J. E., Kulpa, J. E., and Lee, W. M. (2006) Regulation of acetyl-CoA carboxylase. *Biochem. Soc. Trans. 34*, 223–227.

Chen, D. C., Beckerich, J. M., and Gaillardin, C. (1997) One-step transformation of the dimorphic yeast *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol. 48*, 232–235.

Chuang, L.-T., Chen, D.-C., Nicaud, J.-M., Madzak, C., Chen, Y.-H., and Huang, Y.-S. (2010) Co-expression of heterologous desaturase genes in *Yarrowia lipolytica*. *New Biotechnology 27*, 277–282.

Courchesne, N. M. D., Parisien, A., Wang, B., and Lan, C. Q. (2009) Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *J. Biotechnol. 141*, 31–41.

Cui, W., Wang, Q., Zhang, F., Zhang, S.-C., Chi, Z.-M., and Madzak, C. (2011) Direct conversion of inulin into single cell protein by the engineered *Yarrowia lipolytica* carrying inulinase gene. *Process Biochemistry 46*, 1442–1448.

Dulermo, T., and Nicaud, J.-M. (2011) Involvement of the G3P shuttle and $\beta$-oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. *Metab. Eng. 13*, 482–491.

Folch, J., Lees, M., and Sloane-Stanley, G. H. (1957) A simple method for the isolation and purification of total lipids from animal tissues. *The Journal of Biological Chemistry 226*, 497–509.

Hill, J., Nelson, E., Tilman, D., Polasky, S., and Tiffany, D. (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences 103*, 11206.

Huo, Y. X., Cho, K. M., Rivera, J. G. L., Monte, E., Shen, C. R., Yan, Y., and Liao, J. C. (2011) Conversion of proteins into biofuels by engineering nitrogen flux. *Nat. Biotechnol. 29*, 346–351.

Kamiryo, T., Nishikawa, Y., Mishina, M., Terao, M., and Numa, S. (1979) Involvement of long-chain acyl coenzyme A for lipid synthesis in repression of acetyl-coenzyme A carboxylase in Candida lipolytica. *Proceedings of the National Academy of Sciences 76*, 4390.

Kamisaka, Y., Tomita, N., Kimura, K., Kainou, K., and Uemura, H. (2007) DGA1 (diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the δsnf2 disruptant of *Saccharomyces cerevisiae*. *Biochemal Journal 408*, 61–68.

Karatay, S. E., and Dönmez, G. (2010) Improving the lipid accumulation properties of the yeast cells for biodiesel production using molasses. *Bioresour. Technol. 101*, 7988–7990.

Keasling, J. D. (2010) Manufacturing Molecules Through Metabolic Engineering. *Science 330*, 1355–1358.

Klaus, D., Ohlrogge, J. B., Neuhaus, H. E., and Dörmann, P. (2004) Increased fatty acid production in potato by engineering of acetyl-CoA carboxylase. *Planta 219*, 389–396.

Liu, X., Sheng, J., and Curtiss Iii, R. (2011) Fatty acid production in genetically modified cyanobacteria. *Proceedings of the National Academy of Sciences 108*, 6899–6904.

Morin, N., Cescut, J., Beopoulos, A., Lelandais, G., Le Berre, V., Uribelarrea, J.-L., Molina-Jouve, C., and Nicaud, J.-M. (2011) Transcriptomic Analyses during the Transition from Biomass Production to Lipid Accumulation in the Oleaginous Yeast *Yarrowia lipolytica*. *PLoS ONE 6*, e27966.

Oelkers, P., Cromley, D., Padamsee, M., Billheimer, J. T., and Sturley, S. L. (2002) The DGA1 gene determines a second triglyceride synthetic pathway in yeast. *J. Biol. Chem. 277*, 8877.

Ohlrogge, J. B., and Jaworski, J. G. (1997) Regulation of fatty acid synthesis. *Annu. Rev. Plant Biol. 48*, 109–136.

Papanikolaou, S., and Aggelis, G. (2003) Modeling lipid accumulation and degradation in *Yarrowia lipolytica* cultivated on industrial fats. *Curr. Microbiol. 46*, 398–402.

Papanikolaou, S., and Aggelis, G. (2002) Lipid production by *Yarrowia lipolytica* growing on industrial glycerol in a single-stage continuous culture. *Bioresour. Technol. 82*, 43–49.

Papanikolaou, S., Chevalot, I., Komaitis, M., Marc, I., and Aggelis, G. (2002) Single cell oil production by *Yarrowia lipolytica* growing on an industrial derivative of animal fat in batch cultures. *Appl. Microbiol. Biotechnol. 58*, 308–312.

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., and Aggelis, G. (2006) Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by *Yarrowia lipolytica*. *Curr. Microbiol. 52*, 134–142.

Papanikolaou, S., Muniglia, L., Chevalot, I., Aggelis, G., and Marc, I. (2003) Accumulation of a cocoa-butter-like lipid by *Yarrowia lipolytica* cultivated on agro-industrial residues. *Curr. Microbiol. 46*, 124–130.

Papanikolaou, S., Muniglia, L., Chevalot, I., Aggelis, G., and Marc, I. (2002) *Yarrowia lipolytica* as a potential producer of citric acid from raw glycerol. *Journal of Applied Microbiology 92*, 737–744.

Ratledge, C. In *Single Cell Oil*; Moreton, R., Ed.; Longman Scientific & Technical, 1988; Chapter Biochemistry, stoichiometry, substrate and economics, pp 33–70.

Roesler, K., Shintani, D., Savage, L., Boddupalli, S., and Ohlrogge, J. (1997) Targeting of the Arabidopsis Homomeric Acetyl-Coenzyme A Carboxylase to Plastids of Rapeseeds. *Plant Physiol. 113*, 75–81.

Ruenwai, R., Cheevadhanarak, S., and Laoteng, K. (2009) Overexpression of Acetyl-CoA Carboxylase Gene of Mucor rouxii Enhanced Fatty Acid Content in Hansenula polymorpha. *Mol. Biotechnol. 42*, 327–332.

Sambrook, J., and Russell, D. W. *Molecular cloning: a laboratory manual*; CSHL press, 2001; Vol. 2.

Scioli, C., and Vollaro, L. (1997) The use of *Yarrowia lipolytica* to reduce pollution in olive mill wastewaters. *Water Res. 31*, 2520–2524.

Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., and Liao, J. C. (2011) Driving Forces Enable High-Titer Anaerobic 1-Butanol Synthesis in Escherichia coli. *Appl. Environ. Microbiol. 77*, 2905–2915.

Stephanopoulos, G. (2007) Challenges in engineering microbes for biofuels production. *Science 315*, 801.

Tai, M., and Stephanopoulos, G. (2012) Metabolic engineering: enabling technology for biofuels production. *WIREs Energy and Environment* doi: 10.1002/wene.5.

Tsigie, Y. A., Wang, C.-Y., Truong, C.-T., and Ju, Y.-H. (2011) Lipid production from *Yarrowia lipolytica* Po1g grown in sugarcane bagasse hydrolysate. *Bioresour. Technol. 102*, 9216–9222.

Zhang, H., Damude, H. G., and Yadav, N. S. (2011) Three diacylglycerol acyltransferases contribute to oil biosynthesis and normal growth in *Yarrowia lipolytica. Yeast 1*, 25–38.

Zhao, C. H., Cui, W., Liu, X. Y., Chi, Z. M., and Madzak, C. (2010) Expression of inulinase gene in the oleaginous yeast *Yarrowia lipolytica* and single cell oil production from inulin-containing materials. *Metab. Eng. 12*, 510–517.

Zhao, S., and Fernald, R. D. (2005) Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. *J. Comput. Biol. 12*, 1047–1064.

# Chapter 7

# Combinatorial Engineering of Lipid Biosynthesis

## 7.1 Introduction

The study of cellular metabolism can be elucidated using metabolic engineering. Metabolic engineering is the use of recombinant DNA technologies to manipulate metabolic pathways in organisms (Bailey, 1991). Through manipulation and engineering of specific metabolic networks, controlling factors and rate-limiting steps can be identified and elaborated. Building upon existing knowledge and tools for a specific organism or pathway, one can evaluate how novel perturbations can be used to redirect and control the generation of desired products.

Lipid biosynthesis is an excellent pathway for study using for metabolic engineering, having wide applications ranging from health, cancer and medicine, to biochemicals and biofuels production (Kohlwein and Petschnigg, 2007; Beopoulos et al., 2011; Courchesne et al., 2009). The physiology, enzymology, and metabolism for lipid biosynthesis in a wide range of organisms, from bacteria to humans, have been extensively studied, forming a strong knowledge base for both comparative and exploratory analysis (Kurat et al., 2006; Ohlrogge and Jaworski, 1997). Lipid metabolism plays an integral role in numerous aspects of cell physiology, from cell growth and proliferation to energy storage and metabolism (Kohlwein and Petschnigg, 2007; Tehlivets et al., 2007). To utilize these pathways for both medicinal and industrial purposes, it is important to understand which perturbations have the greatest impact on the overall process.

The oleaginous yeast *Yarrowia lipolytica* stands as an excellent model organism to study lipid metabolism. As an oleaginous yeast, *Y. lipolytica* can naturally accumulate up to 36% lipids in carbon rich environments(Beopoulos et al., 2009). These lipids are stored in the form of triacylglycerides (TAG) in lipid bodies. It is one of the most extensively studied 'non-conventional' yeast species, with a sequenced genome and a range of genetic tools available (Barth and Gaillardin, 1997). It has been used in a number of industrial applications and has been viewed as a model organism for protein secretion, hydrophobic substrate utilization, lipid metabolism, and mitochondrial respiration (Beckerich et al., 1998; Coelho et al., 2010; Beopoulos et al., 2009; Kerscher et al., 2002). While *Y. lipolytica* naturally accumulates large

quantities of lipids, a number of engineering efforts have been successful in further increasing or otherwise improving its lipid accumulation characteristics (Dulermo and Nicaud, 2011; Beopoulos et al., 2008; Chuang et al., 2010; Zhang et al., 2011). However the number and variety of genetic manipulations examined has remained relatively limited towards this end and the potential for *Y. lipolytica* as a platform for lipid overproduction remains relatively unexplored.

A number of interesting gene targets have been linked to lipid accumulation through a variety of approaches and strategies (Courchesne et al., 2009). Acetyl-coA carboxylase (ACC) is generally known as the rate-limiting step in fatty biosynthesis, controlling the flux entering the pathway. It is responsible for producing malonyl-coA, which can be utilized in fatty acid elongation. ACC utilizes cytosolic acetyl-coA as its main metabolic precursor. The enzyme that supplies cytosolic acetyl-coA in most eukaryotes is ATP citrate lyase (ACL). ACL cleaves citrate, which has been shuttled out of the mitochondria as a product of the TCA cycle, to form acetyl-coA and oxaloacetate. After fatty acid production is completed with the fatty acid synthase complex, acyl-coA molecules can be further manipulated through elongation and desaturation at the endoplasmic reticulum. These processes help modify the chemical properties of the acyl-coA chains to facilitate storage or utilization in other metabolic pathways. Enzymes such as $\Delta$9-desaturase (D9) convert stearyl-coA molecules into oleoyl-coA molecules, which seem to be very important in both lipid regulation and metabolism (Dobrzyn and Ntambi, 2005). The final step in lipid assembly and storage is the conversion of diacylglycerol (DAG) into TAG via the enzyme diacylglycerol acyltransferase (DGA). This step occurs at both the endoplasmic reticulum and on the surface of lipid bodies, with the latter establishing a dynamic equilibrium of TAG assembly and degradation depending on the energy needs of the organism (Athenstaedt et al., 2006). In *Y. lipolytica,* a number of DGA genes have been identified that perform this function (Zhang et al., 2011).

These enzymatic steps exhibit an interesting relationship to lipid accumulation. ACC controls flux entering lipid synthesis, and overexpression of ACC in the bacteria *Escherichia*

*coli* resulted in 6-fold increase in fatty acid synthesis (Davis et al., 2000). The citrate shuttle, which is under control of ACL, is differentially observed in oleaginous fungi compared to non-oleaginous fungi, and is speculated as a necessary pathway for high flux into the lipid biosynthesis pathway (Boulton and Ratledge, 1981; Vorapreeda et al., 2012). It is also thought that deactivation of ACL leads to citrate accumulation and secretion, an undesirable phenomenon in lipid production (Papanikolaou and Aggelis, 2002; Papanikolaou et al., 2002). D9 has been implicated in cancer metabolism, being upregulated in mammalian tumor cells. It is potentially a strong positive regulator of lipogenesis and facilitates the lipid production necessary for rapid growth found in cancer cells (Ntambi and Miyazaki, 2004; Hulver et al., 2005; Dobrzyn and Ntambi, 2005). DGA is the final committed step for lipid storage, and overexpression of DGA in a *S. cerevisiae* Δsnf2 mutant resulted in dramatic increases in lipid accumulation (Kamisaka et al., 2007). While these results have produced interesting results and implications, analysis of their contributions within a single model organism can allow us to systematically identify how they can contribute and cooperate to achieve the increased lipid production.

Here we look at the impact of several important genes involved in lipid biosynthesis and explore their contributions towards increasing lipid accumulation in the oleaginous yeast *Y. lipolytica*. By overexpression of gene targets, both individually and in combination, we can explore how genes can positively impact flux through lipid biosynthesis pathway. Furthermore we investigate the lipid production performance of two candidate strains to elucidate the importance of balanced metabolic flux within the cell to achieve high productivity.

## 7.2 Materials and Methods

### 7.2.1 Yeast strains, growth, and culture conditions

The *Y. lipolytica* strains used in this study were derived from the wild-type *Y. lipolytica* W29 strain (ATCC20460). The auxotrophic Po1g (Leu-) used in all transformations was

Table 7.1: Survey of combinatorial gene targets. List of all strains examined for lipid accumulation, labeled with the presence of overexpression cassettes (+) for the four targets: acetyl-coA carboxylase (ACC), diacylglycerol acyltransferase (DGA), ATP: citrate lyase (ACL12), and △9-desaturase (D9).

| Plasmid | ACC | DGA | ACL12 | D9 | Number of Targets |
|---|---|---|---|---|---|
| MTYL038 | - | - | - | - | 0 |
| MTYL040 | + | - | - | - | 1 |
| MTYL053 | - | + | - | - | 1 |
| MTYL050 | - | - | + | - | 1 |
| MTYL061 | - | - | - | + | 1 |
| MTYL065 | + | + | - | - | 2 |
| MTYL078 | + | - | + | - | 2 |
| MTYL066 | + | - | - | + | 2 |
| MTYL069 | - | + | - | + | 2 |
| MTYL073 | + | + | - | + | 3 |
| MTYL079 | + | - | + | + | 3 |
| MTYL088 | + | + | + | - | 3 |
| MTYL089 | + | + | + | + | 4 |

obtained from Yeastern Biotech Company (Taipei, Taiwan). All strains used in this study are listed in Table 7.2. Constructed plasmids were linearized with SacII and chromosomally integrated into Po1g according to the one-step lithium acetate transformation method described by Chen et al. (1997). MTYL transformants were named after the numbering of their corresponding integrated plasmids, with the exception of MTYL088 and MTYL089. For construction of strains MTYL088 and MTYL089, strains MTYL078 and MTYL079 underwent two additional rounds of transformation: (1) transformed with URA KO cassette on selective 5-FOA to knock out endogenous URA (as described in Chapter 4 on page 59); (2) transformed with PMT092 linearized with SacII. Transformants were plated on selective media and verified by PCR of prepared genomic DNA. Verified transformants were then stored as frozen glycerol stocks at -80°C and on selective YNB plates at 4°C. A summary of the designed overexpression of genes for each strain is described in Table 7.1.

Media and growth conditions for *Escherichia coli* have been previously described by Sambrook et al.(2001), and those for *Y. lipolytica* have been described by Barth and Gaillardin (1997). Rich medium (YPD) was prepared with 20 g/L Bacto peptone (Difco Laboratories,

Detroit, MI), 10 g/L yeast extract (Difco), 20 g/L glucose (Sigma-Aldrich, St. Louis, MO). YNB medium was made with 1.7 g/L yeast nitrogen base (without amino acids) (Difco), 0.69 g/L CSM-Leu (MP Biomedicals, Solon, OH), and 20 g/L glucose. Selective YNB plates contained 1.7 g/L yeast nitrogen base (without amino acids), 0.69 g/L CSM-Leu, 20 g/L glucose, and 15 g/L Bacto agar (Difco).

Shake flask experiments were carried out using the following medium: 1.7 g/L yeast nitrogen base (without amino acids), 1.5 g/L yeast extract, and 50 g/L glucose. From frozen stocks, precultures were inoculated into YNB medium (5 mL in Falcon tube, 200 rpm, 28°C, 24 hr). Overnight cultures were inoculated into 50 mL of media in 250 mL Erlenmeyer shake flask to an optical density ($A_{600}$) of 0.05 and allowed to incubate for 100 hours (200 rpm, 28°C), after which biomass, sugar content, and lipid content were taken and analyzed.

Bioreactor scale fermentation was carried out in a 2-liter baffled stirred-tank bioreactor. The medium used contained 1.7 g/L yeast nitrogen base (without amino acids and ammonium sulfate), 2 g/L ammonium sulfate, 1 g/L yeast extract, and 90 g/L glucose. From a selective plate, an initial preculture was inoculated into YPD medium (40 mL in 250 mL Erlenmeyer flask, 200 rpm, 28°C, 24 hr). Exponentially growing cells from the overnight preculture were transferred into the bioreactor to an optical density ($A_{600}$) of 0.1 in the 2-L reactor (2.5 vvm aeration, pH 6.8, 28°C, 250 rpm agitation). Time point samples were stored at -20°C for subsequent lipid analysis. Sugar organic acid content was determined by HPLC. Biomass was determined by determined gravimetrically from samples washed and dried at 60°C for two nights.

### 7.2.2 Genetic Techniques

Standard molecular genetic techniques were used throughout this study (Sambrook and Russell, 2001). Restriction enzymes and Phusion High-Fidelity DNA polymerase used in cloning were obtained from New England Biolabs (Ipswich, MA). Genomic DNA from yeast transformants was prepared using Yeastar Genomic DNA kit (Zymo Research, Irvine, CA).

All constructed plasmids were verified by sequencing. PCR products and DNA fragments were purified with PCR Purification Kit or QIAEX II kit (Qiagen, Valencia, CA). Plasmids used are described in Table 7.2. Primers used are described in Table 7.3.

### 7.2.3 Combinatorial plasmid construction

The construction of plasmids pMT010, pMT015, pMT038, pMT040, pMT013 and pMT065 were described previously in Chapter 6 on page 93, but are similar to methods described here. Plasmid pMT047 expressing ATP:citrate lyase subunit 1 was constructed by amplifying the ACL1 gene from *Y. lipolytica* Po1g genomic DNA (Accession Number: XM_504787) and inserting it into the MluI and NsiI sites of pMT010 using primers MT252 and MT253. Plasmid pMT049 expressing ATP:citrate lyase subunit 2 was likewise constructed by amplifying the ACL2 gene from *Y. lipolytica* Po1g genomic DNA (Accession Number: XM_503231) and inserting it into the MluI and NsiI sites of pMT010 using primers MT254 and MT255. Plasmid pMT061 expressing Delta-9 fatty acid desaturase (D9) was amplified from Po1g genomic DNA (Accession Number: XM_501496) and inserted into pINA1269 under control of the php4d promoter using the restriction sites PmlI and BamHI with primers MT283 and MT284.

To produce plasmids that express multiple genes (pMT050, pMT065, pMT066, pMT073, pMT074, pMT075, pMT078, pMT079), a promoter-gene-terminator cassette was amplified from one plasmid using primers MT220 and MT265. This was then digested with DpnI and AseI and inserted into a second plasmid that was digested with NruI and AseI resulting in tandem gene construct pMT065. The AseI restriction site was selected to facilitate selection, as it resides within the ampicillin resistance marker. Because NruI is a blunt end restriction site, ligation of the insertion does not increase the total number of NruI sites, thus enabling iterative insertions using the same process. The overall sequence and scheme used for the combinatorial construction of these plasmids is described in Figure 7.2.1.

Table 7.2: Strains and plasmids used in this study

| Strains (host strain) | Genotype or plasmid (selective marker) | Source |
|---|---|---|
| *E. coli* | | |
| NEB 10$\beta$ | araD139 $\Delta$(ara-leu)7697 fhuA lacX74 galK ($\Phi$80 $\Delta$(lacZ)M15) mcrA galU recA1 endA1 nupG rpsL (StrR) $\Delta$(mrr-hsdRMS-mcrBC) | New England Biolabs |
| pINA1269 | JMP62-LEU | Yeastern |
| JMP-URA | JMP62-URA | Nicaud et al.(2002) |
| pMT010 | pINA1269 php4d::TEF (LEU) | Chapter 6 |
| pMT015 | pINA1269 php4d::TEFin (LEU) | Chapter 6 |
| pMT038 | YTEF-LacZ (LEU) | Chapter 6 |
| pMT040 | YLEX-ACC (LEU) | Chapter 6 |
| pMT013 | YTEF-ACC1 (LEU) | Chapter 6 |
| pMT047 | YTEF-ACL1 (LEU) | This work |
| pMT049 | YTEF-ACL2 (LEU) | This work |
| pMT050 | YTEF-ACL1 + YTEF-ACL2 (LEU) | This work |
| pMT053 | YTEFin-DGA (LEU) | Chapter 6 |
| pMT061 | php4d-D9 (LEU) | This work |
| pMT065 | YLEX-ACC + YTEFin-DGA (LEU) | Chapter 6 |
| pMT066 | YTEF-ACC + YLEX-D9 (LEU) | This work |
| pMT073 | YTEF-ACC + YLEX-D9 + YTEFin-DGA (LEU) | This work |
| pMT074 | YTEF-ACC + YTEF-ACL1 (LEU) | This work |
| pMT075 | YTEF-ACC + YLEX-D9 + YTEF-ACL1 (LEU) | This work |
| pMT078 | YTEF-ACC + YTEF-ACL1 + YTEF-ACL2 (LEU) | This work |
| pMT079 | YTEF-ACC + YLEX-D9 + YTEF-ACL1 + YTEF-ACL2 (LEU) | This work |
| pMT092 | YTEFin-DGA (URA) | Chapter 4 |
| *Y. lipolytica* | | |
| Po1g | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 | Yeastern |
| MTYL038 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-LacZ | This work |
| MTYL040 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 hp4d-ACC1-LEU2 | Chapter 6 |
| MTYL050 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-ACL1 + TEF-ACL2 | This work |
| MTYL053 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEFin-DGA1 | Chapter 6 |
| MTYL065 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 hp4d-ACC1 + TEFin-DGA1 | Chapter 6 |
| MTYL066 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-ACC + hp4d-D9 | This work |
| MTYL073 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-ACC + hp4d-D9 + TEFin-DGA1 | This work |
| MTYL078 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-ACC + TEF-ACL1 + TEF-ACL2 | This work |
| MTYL079 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-ACC + hp4d-D9 + TEF-ACL1 + TEF-ACL2 | This work |
| MTYL088 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-ACC + TEF-ACL1 + TEF-ACL2 + TEFin-DGA | This work |
| MTYL089 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-ACC + hp4d-D9 + TEF-ACL1 + TEF-ACL2 + TEFin-DGA | This work |

Table 7.3: Primers used in this study. Relevant restriction sites are in bold.

| Primer | Description | Sequence |
|---|---|---|
| *PCR* | | |
| MT252 | ACL1 | *CTTACAA**CGCGT**ATGTCTGCCAACGAGAACATCTCC* |
| MT253 | ACL1 | *CTAGGT**ATGCAT**CTATGATCGAGTCTTGGCCTTGGA* |
| MT254 | ACL2 | *CTTACAA**CGCGT**ATGTCAGCGAAATCCATTCACGA* |
| MT255 | ACL2 | *CTAGGT**ATGCAT**TTAAACTCCGAGAGGAGTGGAAGC* |
| MT283 | D9 | *AATGGTGAAAAACGTGGACCAAGTG* |
| MT284 | D9 | *CTCACA**GGATCC**CTAAGCAGCCATGCCAGACATACC* |
| MT220 | Cassette | *CCCGGCAACA**ATTAAT**AGACTGGAT* |
| MT265 | Cassette | *TTCGGACACGGGCATCTCAC* |
| *RT-PCR* | | |
| MTR001 | Actin | *TCCAGGCCGTCCTCTCCC* |
| MTR002 | Actin | *GGCCAGCCATATCGAGTCGCA* |
| MTR031 | DGA | *AACGGAGGAGTGGTCAAGCGA* |
| MTR032 | DGA | *TTATGGGGAAGTAGCGGCCAA* |
| MTR033 | D9 | *GACCGACTCCAACAAGGACCCTT* |
| MTR034 | D9 | *GGGTTGGGCACAAGCAGCAT* |
| MTR041 | ACC | *GCTTCTCACGGAGGTCATACAGT* |
| MTR042 | ACC | *CTGCGGCAATACCGTTGTTAG* |
| MTR068 | ACL1 | *CCTCCTTGGTGAGGTTGGTGGT* |
| MTR069 | ACL1 | *GCGACGATGGGCTTCTTGATC* |
| MTR070 | ACL2 | *CCTTCAAGGGCATCATCCGG* |
| MTR071 | ACL2 | *CGCCTCGTCGCACGTAAATCT* |

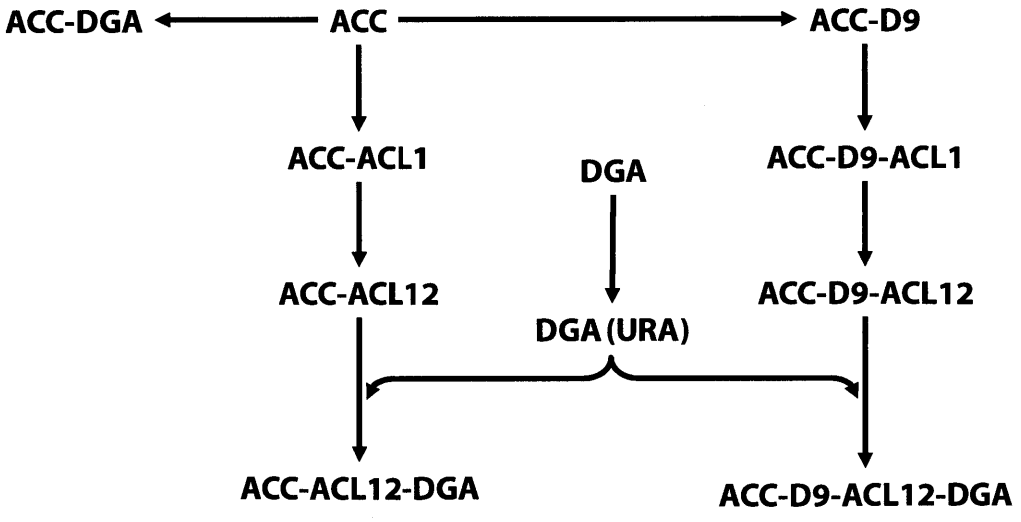


Figure 7.2.1: Combinatorial expression construction scheme. Cloning path and strategy for construction of plasmids containing a combination of lipid accumulation gene targets. These plasmids were then transformed into *Y. lipolytica* for study of lipid accumulation.

### 7.2.4 RNA isolation and transcript quantification

Shake flask cultures grown for 42 hrs were collected and centrifuged for 5 min at 10,000g. Each pellet was resuspended in 1.0 ml of Trizol reagent (Invitrogen) and 100 $\mu$L of acid-washed glass beads were added (Sigma-Aldrich). Tubes were vortexed for 15 min at 4°C for cell lysis to occur. The tubes were then centrifuged for 10 min at 12,000g at 4°C and the supernatant was collected in a fresh 2-mL tube. 200 $\mu$L chloroform was then added and tubes were shaken by hand for 10 seconds. The tubes were again centrifuged for 10 min at 12,000g at 4°C. 400 $\mu$L of the upper aqueous phase was transferred to a new tube, and an equal volume of phenol-chloroform-isoamyl alcohol (pH 4.7) (Ambion, Austin, TX) was added. Tubes were again shaken by hand for 10 seconds and centrifuged for 10 min at 12,000g at 4°C. 250 $\mu$L of the upper phase was transferred to a new tube with an equal volume of cold ethanol and 1/10th volume sodium acetate (pH 5.2). Tubes were chilled at -20°C for thirty minutes to promote precipitation. Tubes were then centrifuged for 5 min at 12,000g, washed twice with 70% ethanol, dried in a 60°C oven and finally resuspended in RNase free water. RNA quantity was analyzed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and samples were stored in -80°C freezer. qRT-PCR analyses were carried out using iScript One-step RT-PCR Kit with SYBR Green (Bio-Rad, Hercules, CA) using the Bio-Rad iCycler iQ Real-Time PCR Detection System. Fluorescence results were analyzed using Real-time PCR Miner and relative quantification and statistical analysis was determined with REST 2009 (Qiagen) using actin as the reference gene and MTYL038 as the reference strain (Zhao and Fernald, 2005). Primers used for qRT-PCR are given in Table 7.3. Samples were analyzed in quadruplicate.

### 7.2.5 Lipid extraction and quantification

Total lipids were extracted using the procedure by Folch et al (1957). A measured quantity of cell biomass (roughly 1 mg) was suspended in 1 mL of chloroform:methanol (2:1) solution and vortexed for 1 hour. After centrifugation, 500 $\mu$L was transferred to 125 $\mu$L saline

solution. The upper aqueous layer was removed and the bottom layer was evaporated and resuspend in 100 $\mu$L hexane. Samples were then stored at -20°C until transesterification.

Transesterification of total lipid extracts was performed by adding 1 mL 2% (wt/vol) sulfuric acid in methanol to each sample. Samples were then incubated at 60°C for 2 hours. After that the samples were partially evaporated, and the fatty acid methyl esters (FAME) were extracted by adding 1 mL hexane and vortexing for 10 min. 800 $\mu$L of this hexane was then transferred into glass vials for GC analysis.

GC analysis of FAMEs was performed with a Bruker 450-GC instrument equipped with a flame-ionization detector and a capillary column HP-INNOWAX (30 m $\times$ 0.25 mm). The GC oven conditions were as follows: 150°C (1 min), a 10 min ramp to 230°C, hold at 230°C for 2 min. The split ratio was 10:1. Fatty acids were identified and quantified by comparison with commercial FAME standards normalized to methyl tridecanoate (C13:0). Total lipid content was calculated as the sum of total fatty acid contents for five FAMEs: methyl palmitate (C16:0), methyl palmitoleate (C16:1), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2) (Sigma-Aldrich). The addition of tridecanoic acid to the chloroform-methanol extraction fluid was used as the internal standard, which was carried through the entire analysis procedure and transesterified into its methyl ester.

### 7.2.6 Direct transesterification

For routine lipid quantification to determine relative lipid accumulation, a method for direct transesterification of cell biomass was used, adapted from the two-step base-then-acid-catalyzed direct transesterification method developed by Griffiths et al. (2010). A normalized quantity of cell culture was centrifuged and the media supernatant was removed. Samples were then stored in -20 $^{o}$C freezer or directly transesterified. The cell was then resuspended with the addition of 100 $\mu$L of hexane containing 10 mg/mL methyl tridecanoate internal standard. 500 $\mu$L 0.5 N sodium methoxide, prepared by the addition of sodium hydroxide to methanol, was then added to the sample. The sample was then vortexed for 1 hour at room

temperature. Next 40 $\mu$L of sulfuric acid was carefully added to the sample, followed by the addition of 500 $\mu$L of neat hexane. The sample was again vortexed at room temperature for another 30 minutes. 300 $\mu$L of the upper hexane layer was then transferred into a glass vial and run using the GC-FID, under standard operating conditions. Total lipid content was calculated as the sum of total fatty acid content for the five primary FAMEs identified.

## 7.3 Results & Discussion

### 7.3.1 Full survey of combinatorial constructs identifies improved strains with select genes

In order to investigate the contributions and interactions of the gene targets, a survey was performed across various intermediate gene expression combinations, testing for the lipid production capabilities of the transformed strains. Table 7.1 describes the 13 constructed strains and their corresponding gene up-regulation. Lipid measurements were all performed after 100 hrs of culture in order to compare lipid productivity rather than merely lipid accumulation. For industrial purposes, overall productivity is a more important measurement than total lipid content, as a slow growing strain producing high yields might still be less useful than a fast growing, moderately high yielding strain. The results of the complete survey of lipid productivities and yields is depicted in Figure 7.3.1.

Examining strains containing only single gene overexpressions (MTYL040, MTYL053, MTYL050, MTYL061), ACC and DGA have clear improvements in both productivity and yield. ACL and D9 did not have any significant increases in either productivity or yield. These results indicate that for *Y. lipolytica*, ACC and DGA exhibit control over lipid biosynthesis and are rate-limiting steps, while ACL and D9 do not exhibit similar phenomena.

The effects of ACC and DGA are discussed in depth in Chapter 6, but DGA creates driving force by sequestering lipids and depleting acyl-coA intermediates, while ACC increases yields by diverting flux towards lipid synthesis and mobilizing the cytosolic acetyl-coA pool

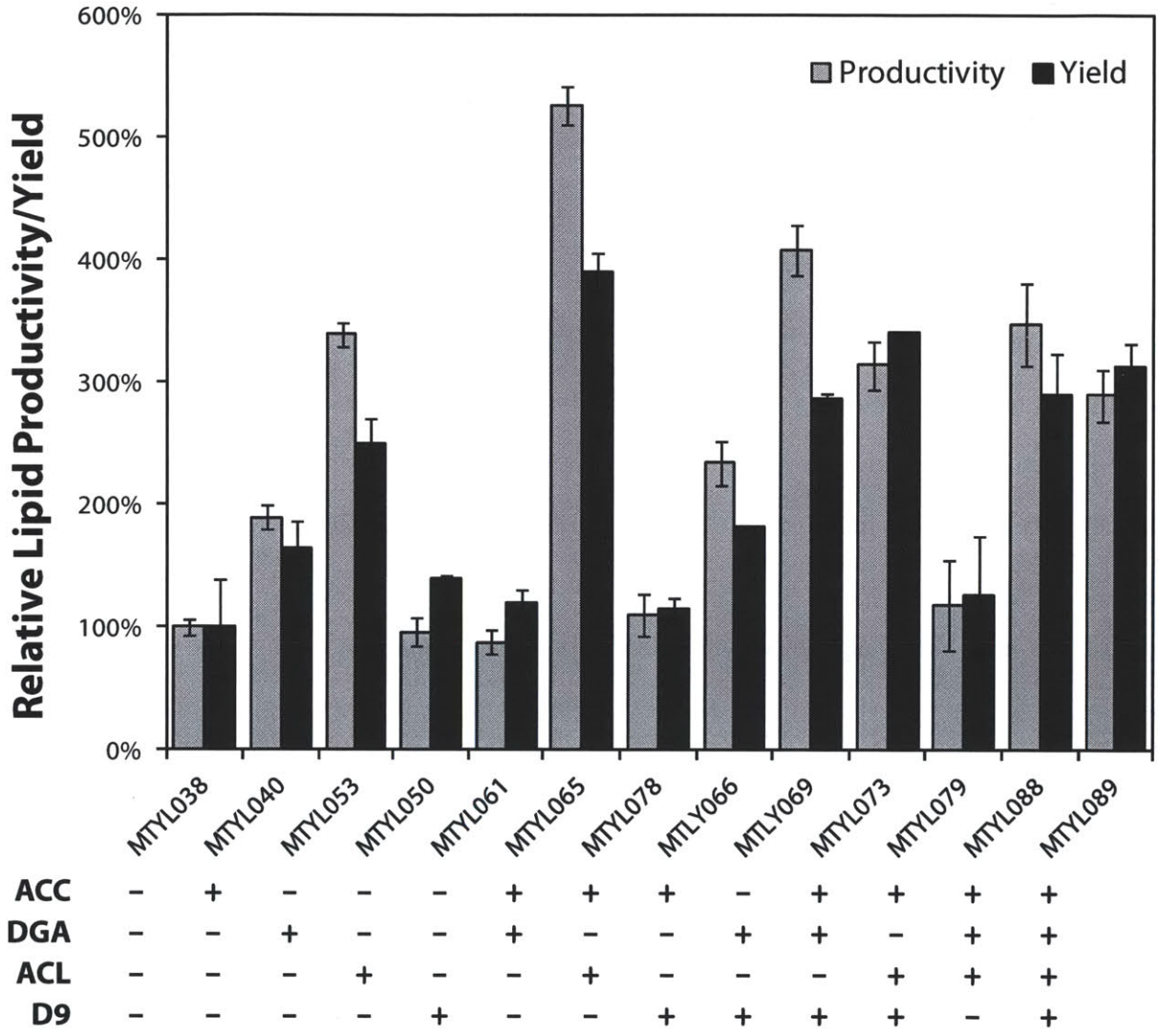


Figure 7.3.1: Relative lipid productivity and yield among *Y. lipolytica* strains expressing combinatorial constructs, as measured by total fatty acid content normalized to the control strain. The presence (+) or absence (-) of a transformed overexpression cassette for the corresponding gene target are indicated below the graph. The productivity (light grey bars) was calculated as relative lipid accumulation within the first 100 hours of culture. Yield calculations were made by dividing lipid accumulation by sugar consumed. The C/N ratio of the media was 20. Results are averaged values across multiple experiments.

more rapidly. When the two genes are combined in strain MTYL065, they produce a synergistic response by establishing a push-and-pull dynamic within the lipid synthesis pathway, with acyl-coA as the balanced intermediate.

When combined with other genes, the gene D9 was able to confer slight benefits to the lipid productivity. For example, MTYL069, overexpressing D9 and DGA, had higher lipid productivities than MTYL053, which contained only DGA overexpression. Likewise, MTYL066, overexpressing ACC and D9, had higher lipid productivities than MTYL040, overexpressing ACC alone. However, MTYL073, overexpressing ACC+D9+DGA, exhibited lower lipid productivity than MTYL065. There were no significant differences between MTYL089 and its D9-lacking variant, MTYL088. While some benefits were observed for accumulation and productivity, the benefits for yield were not significant. The observation that D9 only improves lipid production in combination with other genes seems to suggest that D9 does not have strong regulatory or rate-limiting control over the lipid synthesis process, but its enzymatic action provides favorable conditions magnifying the effect of other genes.

As a membrane-associated enzyme on the lipid body membrane and endoplasmic reticulum, D9 is upregulated during lipid accumulation phases (Morin et al., 2011). Many lipid synthesis enzymes have been found to have the highest specificities for oleate, which is the product of D9 desaturation (Oelkers et al., 2002). This is also demonstrated in the observation that *Y. lipolytica* grows very rapidly on oleate as a carbon source and has extensively been studied growing off of this substrate (Beopoulos et al., 2008; Fickers et al., 2005). Consequently, an increased concentration of oleate, while not specifically driving lipid production or yield, transforms the fatty acid pool to be more rapidly sequestered. This ultimately results in faster rates of lipid accumulation without increases in yield, as increased sequestration will only occur in situations where lipid synthesis has already been upregulated by other manipulations.

In contrast, when combining ACL overexpression with other genes, lipid productivity tended to decrease. MTYL078 and MTYL079 exhibited no significant increases in lipid production, despite overexpressing ACC and/or D9. MTYL088, harboring ACC+ACL+DGA, increased lipid production over control, but did not exhibit any lipid production improvements over MTYL065. These results indicate that while ACL may be affecting the distribution of carbon flux throughout the metabolic network, the overexpression of ACL, whether independently or in combination with other lipogenic improvements, does not significantly promote lipid production and in most cases lowers lipid productivity. This is similar to the observation that while ATP citrate lyase is an enzyme differentially expressed in oleaginous yeast compared to non-oleaginous yeast, the activity of the gene in various organisms has no correlation with the measured oleaginicity (Boulton and Ratledge, 1981).

Another observation from the lipid survey is that expression of DGA along with multiple other targets typically resulted in similar responses, both in productivity and yield. While it is possible that there is some saturation in expression or activity occurring in these constructs, this plateaued response may also have been due to a limitation of the experiment used, as the measured characteristic in this survey was initial overall productivity, rather than stationary phase lipid accumulation or productivity. Furthermore, while strain MTYL065 clearly demonstrated the strongest productivity, the yield was relatively similar to many of the plateaued strains. This suggests that while all of these strains successfully divert flux towards lipids, giving increased yields, MTYL065 is exceptional in its balance of upstream and downstream pathways to achieve both high productivity and yield. These results highlight the importance of balancing perturbations to the metabolic flux network in order to achieve optimal productivities and yields, which is a common theme in metabolic engineering for growth-coupled products (Feist et al., 2010; Tyo et al., 2007).

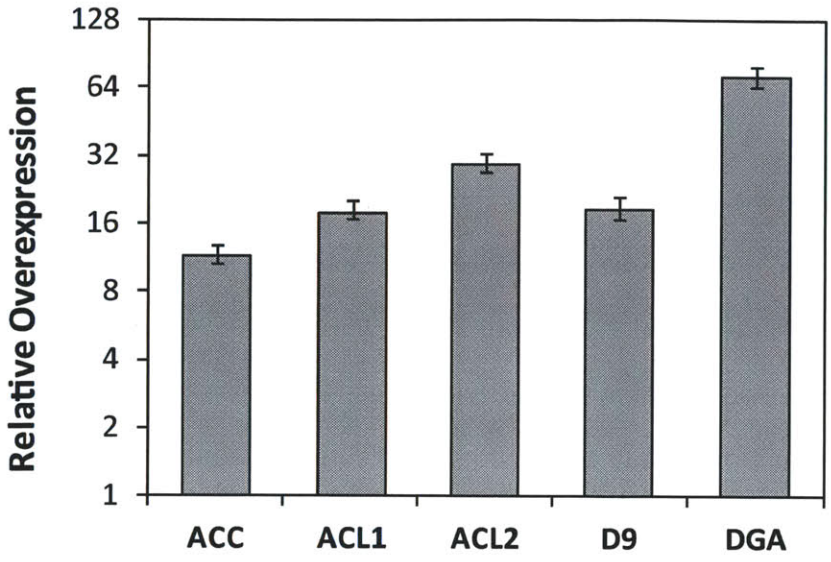


Figure 7.3.2: Transcriptional expression of target genes in the strain MTYL089. Expression is internally normalized to actin expression, and compared against a control (MTYL038) strain and was taken after 66 hours of growth.

### 7.3.2 RT-PCR analysis of full construct shows overexpression in MTYL089

To explore the plateaued region of the lipid survey, the strain MTYL089, overexpressing ACC+D9+ACL12+DGA, was further investigated. The strain was constructed from the transformation of two plasmids, pMT079 and pMT092, into the △LEU and △URA Po1z background strain of *Y. lipolytica*. The use of two plasmids was primarily due to plasmid size considerations, as pMT079 already included four tandem expression cassettes and was 23 kb in length. PCR of genomic DNA confirmed the successful integration of both plasmids into the strain, with confirmation of correct integration of each individual expression cassette. RT-PCR analysis of the completed and verified strain relative to the control strain confirmed proper transcriptional overexpression of all five genes (Figure 7.3.2). The strong overexpression of DGA, which was the only gene under TEFin expression, demonstrates the enhancement characteristics of the spliceosomal intron even with the potential competition from numerous cassettes utilizing the same promoter. ACC, which was under control of

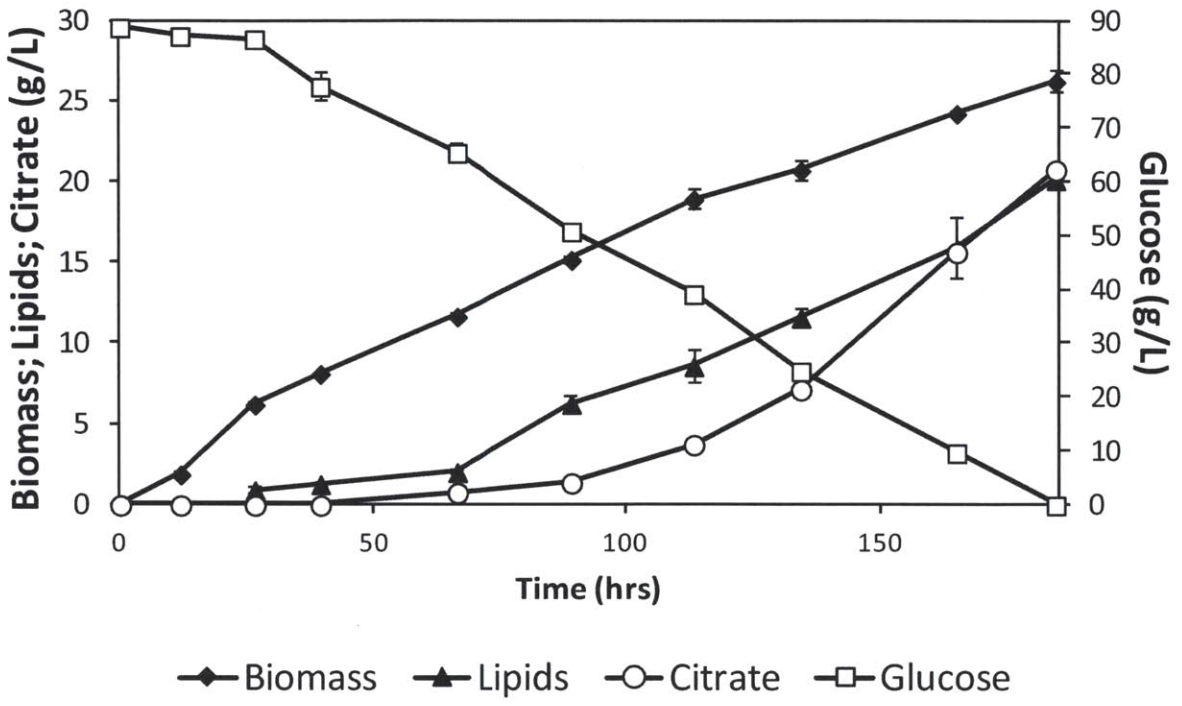


Figure 7.3.3: Batch bioreactor fermentation of strain MTYL089, overexpression ACC, D9, ACL12 and DGA. C/N molar ratio was 100. All sampling was performed in triplicate.

the intronless TEF promoter, showed the lowest expression. The two subunits of ACL, also under control of a TEF promoter, exhibited higher expression than D9, which was under control of an hp4d promoter. Since sampling occurred well after exponential growth phase (where hp4d can exhibit quasi-growth dependent expression(Madzak et al., 2000)), the constitutive expression of the hp4d and TEF all seem to be relatively close to each other. These results show sufficient overexpression of the targeted genes.

### 7.3.3 2-L Fermentation of MTYL089 demonstrates strong lipid accumulation capacity

After verification of gene expression in the strain MTYL089, the lipogenic performance of the strain was tested in a 2-L bioreactor fermentation. The C/N ratio of the media was adjusted to 100 to help promote lipid accumulation. The C/N ratio determines the amount of excess

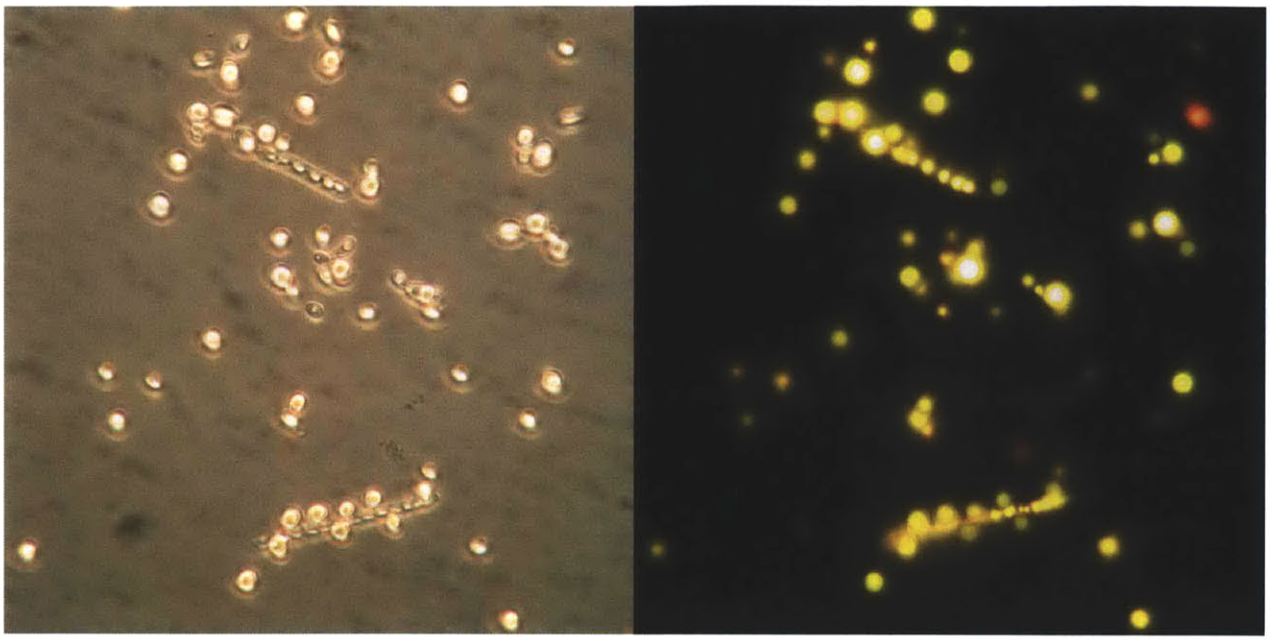

Figure 7.3.4: Microscopy of strain MTYL089 at the end of 2-L fermentation. Normal light microscopy (image left) shows that a majority of the cells are in the yeast form, and contain large vacuoles. Fluorescence microscopy (image right) indicates that these vacuoles are composed of neutral lipids.

carbon available in the fermentation once nitrogen has been depleted, and often requires delicate balancing to optimize lipid production over citrate production(Beopoulos et al., 2009). Figure 7.3.3 shows the time profile for the duration of the batch fermentation. After 185 hrs of fermentation, all 90 g/L of glucose is consumed, yielding 26 g/L of biomass (dry cell weight) and a remarkable 76.8 % lipid content for a productivity of 0.109 g lipids/L/hr. The overall yield of lipids on glucose was 0.227 g lipids/g glucose, which is 70% of the theoretical maximum yield. During the lipid accumulation phase, from 66 to 185 hrs, a maximum in lipid productivity and yield were achieved, at 0.154 g lipids/L/hr and 0.277 g lipids/g glucose, respectively, with the yield increasing to 85% of the maximum theoretical yield. Microscopy of the cell culture, shown in Figure 7.3.4, at the end of the fermentation shows that all the cells contain large vacuoles that occupy virtually the entire volume of the cell. Nile red staining and fluorescence indicates that the vacuoles are predominantly composed of neutral lipids. Even the few hyphal cells, which typically are less productive

(Coelho et al., 2010), exhibit significant lipid accumulation producing multiple lipid bodies along the length of the cell.

Despite the dramatic accumulation of lipids observed in the culture, a number of characteristics were found to be undesirable, particularly in comparison with previous work using MTYL065 (See Chapter 6). Firstly, there was a significant amount of citrate production occurring concomitantly with lipid production beginning at 75 hrs. Citrate is an intermediate of the lipid biosynthesis pathway, utilizing ATP citrate lyase for the enzymatic conversion to oxaloacetate and the lipogenic precursor acetyl-coA. Despite the overexpression of both ACL enzymes in the MTYL089 strain, the accumulation of citrate was still observed. It is possible that the C/N ratio was too high, which has been shown to lead to citrate production instead of lipid production (Beopoulos et al., 2009); however, our results show both citrate and lipids being produced simultaneously. This coupled production of both products differs from a discrete lipid production phase followed by a citrate production phase observed in experiments with strain MTYL065. Furthermore, the C/N ratio was matched to the batch fermentation of MTYL065, which did not show this large amount of citrate production. This indicates that it is more likely that inadequate amounts of ATP generation under the fermentation conditions, combined with high upstream flux into the pathway, led to accumulation of the intracellular citrate pool, ultimately leading to secretion. Additionally, a significantly lower productivity was observed in MTYL089. Since aeration is kept constant throughout the fermentation, oxygen-limited growth is expected in the latter stages of the fermentation. However, the onset of linear growth occurred much earlier in this fermentation than with MTYL065, occurring only after one day when the biomass concentration reached approximately 6 g/L. The earlier onset of linear growth resulted in a longer fermentation time, and thus lower productivity despite the higher lipid content. Because the aeration was also matched to the fermentation conditions of MTYL065, it is likely that the metabolic changes of MTYL089 are putting greater limitations on growth. On the other hand, MTYL089 ex-

hibited better overall lipid yield, 0.227 g/g compared to 0.195 g lipid/g glucose in MTYL065. It also ended the fermentation with a higher titer, 20.2 g/L lipids, compared to 17.6 g/L.

Table 7.4 summarizes the comparison of key performance characteristics between 2-L fermentations of strain MTYL065 and MTYL089. Comparison of the fatty acid profiles (Figure 7.3.5 on the facing page) indicates only slight changes in the distribution of fatty acids between the strains, having stronger preference for the monounsaturated fatty acids palmitoleate and oleate. The additional effects of ACL and D9 overexpression appear to increase the flux towards lipid synthesis, but at a considerable cost to growth rate. Additionally, since there is no matching increase in ATP generation by the cell metabolism, citrate is secreted as a byproduct rather than utilized in the lipid synthesis pathway. Since lipid synthesis and storage can strongly compete with growth for resources, tight regulation is normally necessary to manage this activity (Tehlivets et al., 2007). Overexpression of these four gene targets has unlocked a great deal of this regulation, and as a result we are observing strong competition between prioritization of cellular growth and lipid production. As a common theme in metabolic engineering, maximizing production of a particular product often requires balancing the flux towards the desired product and the overall health and growth of the organism (Chapter 2). The contrasts between MTYL065 and MTYL089 clearly demonstrate this need, with MTYL065 exemplifying more optimized flux through the lipid synthesis pathway.

## 7.4 Conclusion

While the components of the lipid biosynthesis are well-understood, there is still a great deal unknown about how gene perturbations, particularly in combination, affect the capacity and flux through this pathway. By studying the oleaginous yeast *Y. lipolytica*, we are able to utilize a host organism with natural capacity for lipid production to study the extent to which metabolic engineering can improve lipid productivity and yield. We examine four

Table 7.4: Comparison of fermentation characteristics between strains MTYL065 and MTYL089. C/N ratio of both fermentations were 100, performed in 2-L bioreactors. Glucose reactor initially charged with 90 g/L glucose. Further discussion of the MTYL065 strain can be found in Chapter 6 on page 93.

| | MTYL065 | MTYL089 |
|---|---|---|
| Overall Biomass Yield | 0.316 g/g | 0.295 g/g |
| Overall Lipid Yield | 0.195 g/g | 0.227 g/g |
| Overall Productivity | 0.143 g/L/hr | 0.109 g/L/hr |
| Lipid Production Phase | 70 - 120 hrs | 66 - 185 hrs |
| Maximum Lipid Yield | 0.249 g/g | 0.277 g/g |
| Maximum Productivity | 0.178 g/L/hr | 0.154 g/L/hr |
| Final Lipid Titer | 17.6 g/L | 20.2 g/L |

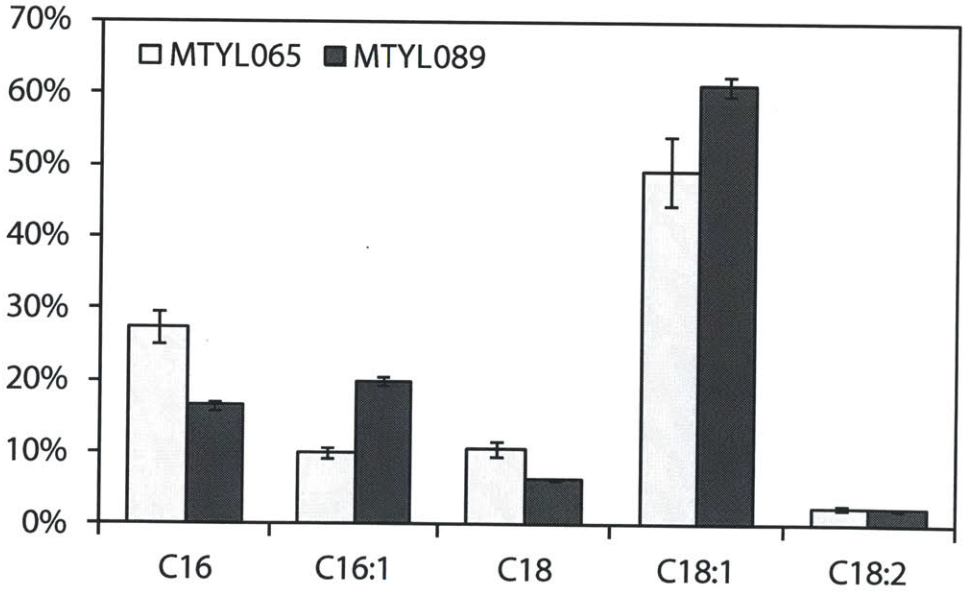


Figure 7.3.5: Comparison of fatty acid profiles between strains MTYL065 and MTYL089. Fatty acid profiles taken from final time point in the respective fermentations, normalizing to the total fatty acid content.

gene targets - ACC, D9, ACL, DGA - and are able to achieve a remarkable 76.8% lipid content in a 2-L bioreactor with a strain carrying all four target overexpressions. By further investigation of these gene targets through combinatorial overexpression, we were able to rank the positive impact of these genes on lipid production, with DGA and ACC being the strong positive contributors, D9 making slight contributions only when combined with other genes, and finally ACL making no significant positive contributions. We were also able to explore possible interactions between the individual effects, identifying the strongest synergistic interaction between ACC and DGA. The production of microbial lipids has a wide range of uses, and has fast gained attention for its utilization in the production of biodiesel. Metabolic engineering of the central pathways of lipid synthesis will be critical in providing success in enabling these future technologies and processes.

# References


Athenstaedt, K., Jolivet, P., Boulard, C., Zivy, M., Negroni, L., Nicaud, J. M., and Chardot, T. (2006) Lipid particle composition of the yeast *Yarrowia lipolytica* depends on the carbon source. *Proteomics 6*, 1450–1459.

Bailey, J. E. (1991) Toward a science of metabolic engineering. *Science 252*, 1668–1675.

Barth, G., and Gaillardin, C. (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica. FEMS Microbiol. Rev. 19*, 219–237.

Beckerich, J. M., Boisramé-Baudevin, A., and Gaillardin, C. (1998) *Yarrowia lipolytica*: a model organism for protein secretion studies. *International Microbiology 1*, 123.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J. L., Molina-Jouve, C., and Nicaud, J. M. (2009) *Yarrowia lipolytica* as a model for bio-oil production. *Progress in Lipid Research 48*, 375–387.

Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M. T., Hapala, I., Papanikolaou, S., Chardot, T., and Nicaud, J. M. (2008) Control of lipid accumulation in the yeast *Yarrowia lipolytica. Appl. Environ. Microbiol. 74*, 7779.

Beopoulos, A., Nicaud, J.-M., and Gaillardin, C. (2011) An overview of lipid metabolism in yeasts and its impact on biotechnological processes. *Appl. Microbiol. Biotechnol. 90*, 1193–1206.

Boulton, C. A., and Ratledge, C. (1981) Correlation of Lipid Accumulation in Yeasts with Possession of ATP: Citrate Lyase. *Journal of General Microbiology 127*, 169–176.

Chen, D. C., Beckerich, J. M., and Gaillardin, C. (1997) One-step transformation of the dimorphic yeast *Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 48*, 232–235.

Chuang, L.-T., Chen, D.-C., Nicaud, J.-M., Madzak, C., Chen, Y.-H., and Huang, Y.-S. (2010) Co-expression of heterologous desaturase genes in *Yarrowia lipolytica*. *New Biotechnology 27*, 277–282.

Coelho, M. A. Z., Amaral, P. F. F., and Belo, I. In *Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology*; Mendez-Vilas, A., Ed.; Formatex Research Center, 2010; Vol. 2; Chapter Yarrowia lipolytica: an industrial workhorse, pp 930–944.

Courchesne, N. M. D., Parisien, A., Wang, B., and Lan, C. Q. (2009) Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *J. Biotechnol. 141*, 31–41.

Davis, M. S., Solbiati, J., and Cronan, J., John E. (2000) Overproduction of Acetyl-CoA Carboxylase Activity Increases the Rate of Fatty Acid Biosynthesis in Escherichia coli. *J. Biol. Chem. 275*, 28593–28598.

Dobrzyn, A., and Ntambi, J. M. (2005) The role of stearoyl-CoA desaturase in the control of metabolism. *Prostaglandins Leukot Essent Fatty Acids 73*, 35–41.

Dulermo, T., and Nicaud, J.-M. (2011) Involvement of the G3P shuttle and $\beta$-oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. *Metab. Eng. 13*, 482–491.

Feist, A. M., Zielinski, D. C., Orth, J. D., Schellenberger, J., Herrgard, M. J., and Palsson, B. O. (2010) Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metab Eng 12*, 173–186.

Fickers, P., Benetti, P. H., Wache, Y., Marty, A., Mauersberger, S., Smit, M. S., and Nicaud, J. M. (2005) Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *FEMS Yeast Res. 5*, 527–543.

Folch, J., Lees, M., and Sloane-Stanley, G. H. (1957) A simple method for the isolation and purification of total lipids from animal tissues. *The Journal of Biological Chemistry 226*, 497–509.

Griffiths, M. J., van Hille, R. P., and Harrison, S. T. L. (2010) Selection of direct transesterification as the preferred method for assay of fatty acid content of microalgae. *Lipids 45*, 1053–1060.

Hulver, M. W., Berggren, J. R., Carper, M. J., Miyazaki, M., Ntambi, J. M., Hoffman, E. P., Thyfault, J. P., Stevens, R., Dohm, G. L., Houmard, J. A., and Muoio, D. M. (2005) Elevated stearoyl-CoA desaturase-1 expression in skeletal muscle contributes to abnormal fatty acid partitioning in obese humans. *Cell Metab 2*, 251–261.

Kamisaka, Y., Tomita, N., Kimura, K., Kainou, K., and Uemura, H. (2007) DGA1 (diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the δsnf2 disruptant of *Saccharomyces cerevisiae*. *Biochemal Journal 408*, 61–68.

Kerscher, S., Dröse, S., Zwicker, K., Zickermann, V., and Brandt, U. (2002) Yarrowia lipolytica, a yeast genetic system to study mitochondrial complex I. *Biochim Biophys Acta 1555*, 83–91.

Kohlwein, S., and Petschnigg, J. (2007) Lipid-induced cell dysfunction and cell death: Lessons from yeast. *Current Hypertension Reports 9*, 455–461.

Kurat, C. F., Natter, K., Petschnigg, J., Wolinski, H., Scheuringer, K., Scholz, H., Zimmermann, R., Leber, R., Zechner, R., and Kohlwein, S. D. (2006) Obese Yeast: Triglyceride Lipolysis Is Functionally Conserved from Mammals to Yeast. *J. Biol. Chem. 281*, 491–500.

Madzak, C., Tréton, B., and Blanchin-Roland, S. (2000) Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *J. Mol. Microbiol. Biotechnol. 2*, 207–216.

Morin, N., Cescut, J., Beopoulos, A., Lelandais, G., Le Berre, V., Uribelarrea, J.-L., Molina-Jouve, C., and Nicaud, J.-M. (2011) Transcriptomic Analyses during the Transition from Biomass Production to Lipid Accumulation in the Oleaginous Yeast *Yarrowia lipolytica*. *PLoS ONE 6*, e27966.

Nicaud, J.-M., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P., and Gaillardin, C. (2002) Protein expression and secretion in the yeast *Yarrowia lipolytica*. *FEMS Yeast Res. 2*, 371–379.

Ntambi, J. M., and Miyazaki, M. (2004) Regulation of stearoyl-CoA desaturases and role in metabolism. *Prog Lipid Res 43*, 91–104.

Oelkers, P., Cromley, D., Padamsee, M., Billheimer, J. T., and Sturley, S. L. (2002) The DGA1 gene determines a second triglyceride synthetic pathway in yeast. *J. Biol. Chem. 277*, 8877.

Ohlrogge, J. B., and Jaworski, J. G. (1997) Regulation of fatty acid synthesis. *Annu. Rev. Plant Biol. 48*, 109–136.

Papanikolaou, S., and Aggelis, G. (2002) Lipid production by *Yarrowia lipolytica* growing on industrial glycerol in a single-stage continuous culture. *Bioresour. Technol. 82*, 43–49.

Papanikolaou, S., Muniglia, L., Chevalot, I., Aggelis, G., and Marc, I. (2002) *Yarrowia lipolytica* as a potential producer of citric acid from raw glycerol. *Journal of Applied Microbiology 92*, 737–744.

Sambrook, J., and Russell, D. W. *Molecular cloning: a laboratory manual*; CSHL press, 2001; Vol. 2.

Tehlivets, O., Scheuringer, K., and Kohlwein, S. D. (2007) Fatty acid synthesis and elongation in yeast. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1771*, 255–270.

Tyo, K. E., Alper, H. S., and Stephanopoulos, G. N. (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol. 25*, 132–137.

Vorapreeda, T., Thammarongtham, C., Cheevadhanarak, S., and Laoteng, K. (2012) Alternative routes of acetyl-CoA synthesis identified by comparative genomic analysis: involvement in the lipid production of oleaginous yeast and fungi. *Microbiology 158*, 217–228.

Zhang, H., Damude, H. G., and Yadav, N. S. (2011) Three diacylglycerol acyltransferases contribute to oil biosynthesis and normal growth in *Yarrowia lipolytica. Yeast 1*, 25–38.

Zhao, S., and Fernald, R. D. (2005) Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. *J. Comput. Biol. 12*, 1047–1064.

# Chapter 8

# Exploring Xylose Utilization in *Yarrowia lipolytica*

## 8.1 Introduction

In the search for improved feedstocks, the push towards cellulosic biofuels is a clear choice. Cellulosic biomass mitigates the need to compete with food crop production; an estimated 1.3+ billion dry tons per year of biomass is potentially available in the US alone (Perlack, 2005). Additionally, cellulosic materials can be more efficiently grown and more stably produced compared to sugar crops. However cellulosic materials are not naturally consumable by most biofuel-producing organisms, and thus cellulose requires pretreatment and hydrolysis to break the material down into monomeric sugar. The resulting hydrolysate can then be used as a sugar rich feedstock.

Since hydrolysis of lignocellulosic biomass results in 20-30% carbohydrates in the form of xylose, utilization of pentose sugars is one of the first steps toward efficiently using cellulosic materials. *Saccharomyces cerevisiae*, the most productive of ethanologenic organisms, cannot ferment xylose; it lacks the ability to convert xylose into xylulose, which can then enter the pentose phosphate pathway (PPP). Transferring the xylose reductase (XR or XYL1) and xylitol dehydrogenase (XDH or XYL2) enzymes from *Scheffersomyces stipitis* (formerly *Pichia stipitis*) has been shown to enable growth of the yeast on xylose for production of ethanol (Jeffries, 2006). The addition of xylulokinase (XK or XYL3) can also be used to further improve utilization, although *S. cerevisiae* already carries an endogenous version of this gene. A secondary pathway, using xylose isomerase (XYLA), can be used to convert xylose into xylulose. Compared to the XR/XDH redox pathway, which uses NADPH and $NAD^+$ cofactors for shuttling of reducing equivalents, the isomerase pathway requires no cofactors. Nonetheless the redox pathway is much more prevalent in nature, and likewise in literature (Jeffries, 2006; Matsushika et al., 2009).

Instead of ethanol production, it may also be advantageous to produce yeast oil for biodiesel from cellulosic feedstocks. As a robust lipid producing organism, *Y. lipolytica* appears to be an attractive platform for the production of cellulosic biodiesel. By leveraging the knowledge and resources developed for xylose metabolic engineering in *S. cerevisiae*,

xylose utilization in *Y. lipolytica* enables the robust production of yeast oils from cellulosic materials. Because theoretical yields of lipid production from xylose are very similar to that of glucose (0.34 g/g compared to 0.32 g/g), the consumption of xylose represents an attractive and worthwhile opportunity in a developing cellulosic biodiesel microbial bioprocess (Ratledge, 1988). Furthermore, *Y. lipolytica* has a very high relative PPP flux (Blank et al., 2005), a phenotype advantageous for growth on xylose since all flux must pass through the PPP. Upregulation of the PPP pathway is a commonly engineered aspect in xylose utilizing *S. cerevisiae* strains (Walfridsson et al., 1995).

Figure 8.1.1 depicts the overall metabolic network for lipid production with the incorporation of the xylose utilization pathway. Xylose enters the cell and can be catabolized either through the redox (XR/XDH) pathway or the isomerase (XYLA) pathway, producing xylulose. It can then enter central metabolism through the non-oxidative pathway of the PPP where it ultimately produces glyceraldehyde-3-phosphate (G3P) and fructose-6-phosphate (F6P). These two products can then enter the rest of central metabolism, going through glycolysis to enter the TCA cycle. Production of lipids occurs normally through the transport of mitochondrial citrate into the cytosol, where it is cleaved by ATP citrate lyase into oxaloacetate and cytosolic acetyl-coA. The acetyl-coA can then enter the fatty acid synthesis pathway through the enzymatic activity of acetyl-coA carboxylase. Acyl-coA generated from the fatty acid synthase complex are transferred to a glycerol-3-phosphate backbone and ultimately sequestered within lipid bodies as triacylglycerol (TAG).

Here we describe the analysis of *Y. lipolytica* for its natural xylose utilization and the metabolic engineering of the organism enabling utilization of xylose for the production of lipids. By incorporation of XR/XDH genes we are able to enable growth on xylose as sole carbon source, and open up opportunities for the production of lipids from cofermentations. Next we study the performance of our engineered strain through the use of cofermentations to analyze for catabolite repression and response, and evaluate the performance of the strain in a scaled-up 2-L bioreactor glycerol-xylose cofermentation with respect to lipid production.

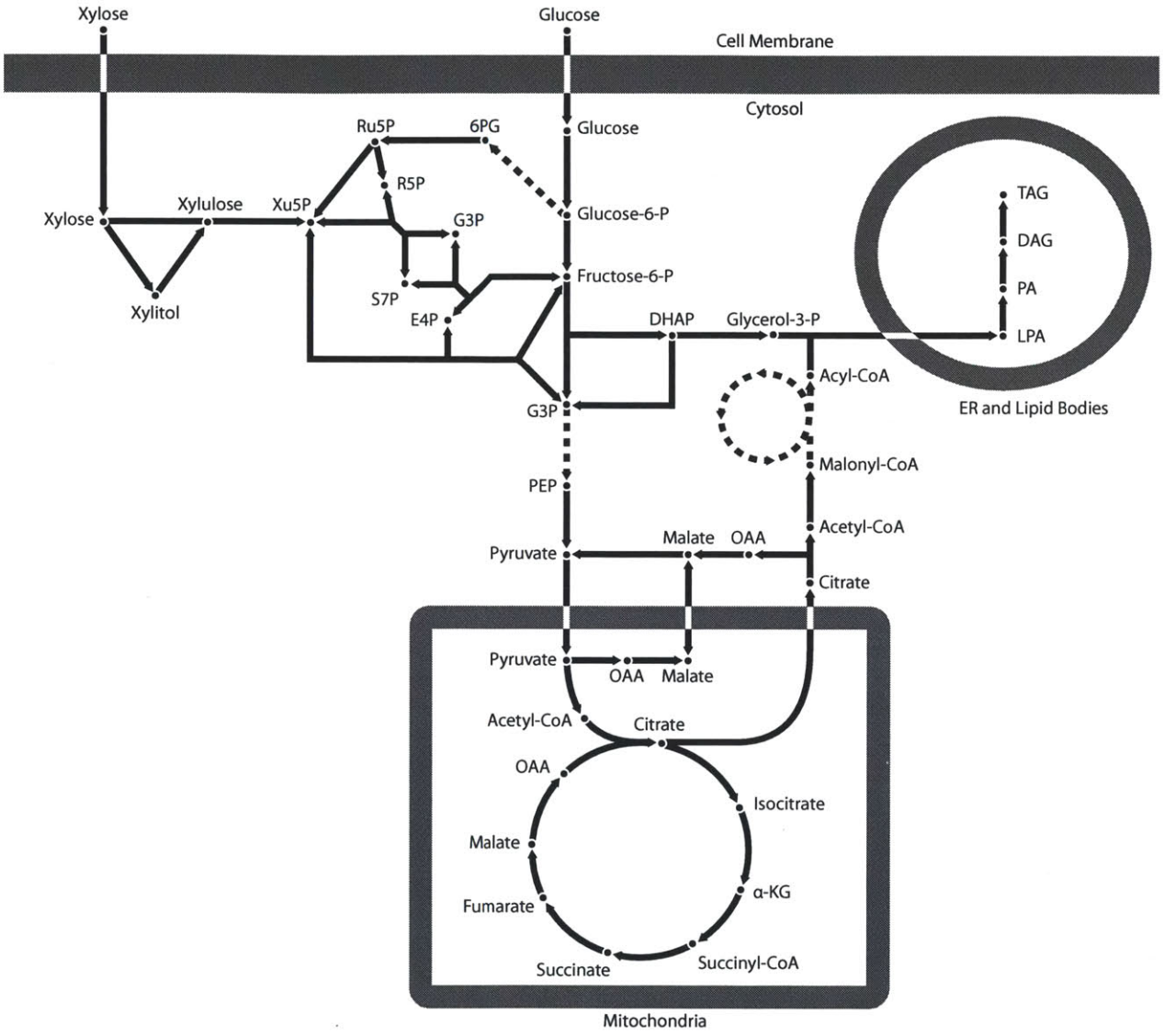


Figure 8.1.1: Metabolic pathway for the conversion of xylose to lipids. Xylose enters the cell and can be converted either through the redox (XYL123) pathway or the isomerase (XYLA) pathway. Next it can enter central metabolism by way of the non-oxidative phase of the pentose phosphate pathway (PPP).

Table 8.1: Strains and plasmids used in this study

| Strains (host strain) | Genotype or plasmid | Source |
|---|---|---|
| *E. coli* | | |
| DH5$\alpha$ | fhuA2 $\Delta$(argF-lacZ)U169 phoA glnV44 $\Phi$80 $\Delta$(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 | Invitrogen |
| pINA1269 | JMP62-LEU | Yeastern |
| pRS426-XYL123 | XYL123 (*P. stipitis*) | Zhou(2011) |
| pMT015 | pINA1269 php4d::pTEFin | Chapter 4 |
| pMT041 | php4d-XYL1 | This work |
| pMT044 | php4d-XYL2 | This work |
| pMT053 | YTEFin-DGA | Chapter 6 |
| pMT059 | TEFin-XYL1 | This work |
| pMT081 | TEFin-XYL1 + hp4d-XYL2 | This work |
| pMT085 | TEFin-XYL1 + hp4d-XYL2 + TEFin-DGA | This work |
| *Y. lipolytica* | | |
| Po1g | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 | Yeastern |
| MTYL038 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEF-LacZ-LEU2 | Chapter 4 |
| MTYL081 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEFin-XYL1 + hp4d-XYL2 | This work |
| MTYL085 | MATa, leu2-270, ura3-302::URA3, xpr2-332, axp-2 TEFin-XYL1 + hp4d-XYL2 + TEFin-DGA | This work |

Finally we perform transcription analysis to observe the respiratory responses of the organism during cofermentation.

## 8.2 Materials and Methods

### 8.2.1 Yeast strains, growth, and culture conditions

The *Y. lipolytica* strains used in this study were derived from the wild-type *Y. lipolytica* W29 strain (ATCC20460). The auxotrophic Po1g (Leu-) used in all transformations was obtained from Yeastern Biotech Company (Taipei, Taiwan). All strains used in this study are listed in Table 8.1. Constructed plasmids were linearized with SacII and chromosomally integrated into Po1g according to the one-step lithium acetate transformation method described by Chen et al. (1997). MTYL transformants were named after the numbering of their corresponding integrated plasmids. Transformants were plated on selective media and verified by PCR of prepared genomic DNA. Verified transformants were then stored as frozen glycerol stocks at -80°C and on selective YNB plates at 4°C.

Table 8.2: Primers used in this study. Relevant restriction sites are in bold.

| Primer | Description | Sequence |
|---|---|---|
| *PCR* | | |
| MT233 | XYL2 | *AATGACTGCTAACCCTTCCTTGGTGT* |
| MT234 | XYL2 | *CTGGTCTAGGT**GGATCC**TTACTCAGGGCCGTCAATGAGAC* |
| MT243 | XYL1 | *AATGCCTTCTATTAAGTTGAACTCTGGTTAC* |
| MT244 | XYL1 | *CTAGGTCTTACT**GGATCC**TTAGACGAAGATAGGAATCTTGTCCCA* |
| MT281 | XYL1 | *TAACCGCAGCATCATCACCATCACCACCCTTCTATTAAGTTGAACTCTGGTTACGAC* |
| MT282 | XYL1 | *CTTACA**GGTACC**TTAGACGAAGATAGGAATCTTGTCCCAG* |
| *RT-PCR* | | |
| MTR001 | Actin | *TCCAGGCCGTCCTCTCCC* |
| MTR002 | Actin | *GGCCAGCCATATCGAGTCGCA* |
| MTR017 | ylXYL1 | *AAGGAGTGGGCTGGATGGA* |
| MTR018 | ylXYL1 | *GGTCTCTCGGGTAGGGATCTTG* |
| MTR019 | ylXYL2 | *ATGGAGGAATCGGCGACTT* |
| MTR020 | ylXYL2 | *ACCACCTCTCCGGCACTTT* |
| MTR031 | DGA | *AACGGAGGAGTGGTCAAGCGA* |
| MTR032 | DGA | *TTATGGGGAAGTAGCGGCCAA* |
| MTR051 | psXYL2 | *CTCCAAGTTGGGTTCCGTTGC* |
| MTR052 | psXYL2 | *GCGACAGCAGCAGCCAAAAGA* |
| MTR053 | psXYL1 | *AGGCTATCGCTGCTAAGCACGG* |
| MTR054 | psXYL1 | *TTTGGAATGATGGCAATGCCTC* |
| MTR055 | ylXYL3 | *CAGCTCAAGGGCATCATTCTGG* |
| MTR056 | ylXYL3 | *TGCGGCAAGTCGTCCTCAAA* |
| MTR060 | IDH1 | *CTTCGAACCGCCTACCTGGCTA* |
| MTR061 | IDH1 | *TGGGCTGGAACATGGTTCGA* |
| MTR064 | ACO1 | *CACCGCTTTCGCCATTGCT* |
| MTR065 | ACO1 | *GGGCTCCTTGAGCTTGAACTCC* |
| MTR066 | PDB1 | *CTGTGGTGTCGTCAACGACTCC* |
| MTR067 | PDB1 | *GCTCAATGGCGTAAGGAGTGG* |
| MTR072 | ICL | *TACTCTCCCGAGGACATTGCC* |
| MTR073 | ICL | *CAGCTTGAAGAGCTTGTCAGCC* |

Media and growth conditions for *Escherichia coli* have been previously described by Sambrook et al.(2001), and those for *Y. lipolytica* have been described by Barth and Gaillardin (1997). Rich medium (YPD) was prepared with 20 g/L Bacto peptone (Difco Laboratories, Detroit, MI), 10 g/L yeast extract (Difco), 20 g/L glucose (Sigma-Aldrich, St. Louis, MO). YNB medium was made with 1.7 g/L yeast nitrogen base (without amino acids) (Difco), 0.69 g/L CSM-Leu (MP Biomedicals, Solon, OH), and 20 g/L glucose. Selective YNB plates contained 1.7 g/L yeast nitrogen base (without amino acids), 0.69 g/L CSM-Leu, 20 g/L glucose, and 15 g/L Bacto agar (Difco).

Shake flask experiments were carried out using the following medium: 1.7 g/L yeast nitrogen base (without amino acids), 1.5 g/L yeast extract, and 50 g/L glucose. From frozen stocks, precultures were inoculated into YNB medium (5 mL in Falcon tube, 200 rpm, 28°C, 24 hr). Overnight cultures grown in YPD were centrifuged, washed, and reinoculated into 50 mL of media in 250 mL Erlenmeyer shake flask (200 rpm, 28°C). OD, biomass and sugar content were taken periodically and analyzed.

For adaptation of strains on xylose, verified transformants were inoculated into shake flasks containing minimal media and 20g/L xylose. The cultures were incubated at 30°C for at least 10 days, waiting for growth to occur, before reinoculation into fresh media. This process was repeated until the final OD of the culture reached at least 20, indicating adaptation to xylose. The culture was then stored as frozen stock in 15% glycerol at -80°C for subsequent use.

Bioreactor scale fermentation was carried out in a 2-liter baffled stirred-tank bioreactor. The medium used contained 1.7 g/L yeast nitrogen base (without amino acids and ammonium sulfate), 2 g/L ammonium sulfate, 1 g/L yeast extract, and 90 g/L glucose. From a selective plate, an initial preculture was inoculated into YPD medium (40 mL in 250 mL Erlenmeyer flask, 200 rpm, 28°C, 24 hr). Exponentially growing cells from the overnight preculture were transferred into the bioreactor to an optical density ($A_{600}$) of 0.1 in the 2-L reactor (2.5 vvm aeration, pH 6.8, 28°C, 250 rpm agitation). Time point samples were stored

at -20°C for subsequent lipid analysis. Sugar organic acid content was determined by HPLC. Biomass was determined by determined gravimetrically from samples washed and dried at 60°C for two nights. Lipid content was analyzed by direct transesterification.

### 8.2.2 Plasmid construction

Standard molecular genetic techniques were used throughout this study (Sambrook and Russell, 2001). Restriction enzymes and Phusion High-Fidelity DNA polymerase used in cloning were obtained from New England Biolabs (Ipswich, MA). Genomic DNA from yeast transformants was prepared using Yeastar Genomic DNA kit (Zymo Research, Irvine, CA). All constructed plasmids were verified by sequencing. PCR products and DNA fragments were purified with PCR Purification Kit or QIAEX II kit (Qiagen, Valencia, CA). Plasmids used are described in Table 8.1. Primers used are described in Table 8.2.

Plasmid pMT041 was constructed by amplifying the xylose reductase gene (XYL1; Accession Number: XM_001385144) from the plasmid pRS426-XYL123 using the primers MT243 and MT244 and inserting it between the PmlI and BamHI sites of pINA1269. Plasmid pMT044 was constructed by amplifying the xylitol dehydrogenase gene (XYL2; Accession Number: XM_001386945) from the plasmid pRS426-XYL123 using the primers MT233 and MT234 and inserting it between the PmlI and BamHI sites of pINA1269. XYL1 and XYL2 are both genes originally from the xylose utilizing yeast, *Scheffersomyces stipitis* (formerly *Pichia stipitis*).

Plasmid pMT059 was constructed by amplifying the XYL1 gene from pMT041 using the primers MT281 and MT282. The amplicon was then inserted into the TEFin expression plasmid, pMT015 between the sites SnaBI and KpnI.

For the expression of multiple genes on a single plasmid, the promoter-gene-terminator cassette can be amplified from a parent vector using primers MT220 and MT265. The cassette can then be inserted into the receiving vector between the restriction sites NruI and AseI, resulting in a tandem gene construct. The AseI restriction site was selected to facilitate

selection, as it resides within the Ampicillin resistance marker of the plasmid. Because NruI is a blunt end restriction site, insertion of the amplicon does not increase the total number of NruI sites that helps facilitate progressive insertions. Plasmid pMT081 was constructed by amplifying the XYL2 cassette from pMT044 and inserting it into the plasmid pMT059, containing XYL1. Plasmid pMT085 was constructed by amplifying the DGA cassette from pMT053 and inserting it into the plasmid pMT081, which contains XYL12.

### 8.2.3 RNA isolation and transcript quantification

Shake flask cultures grown for 42 hrs were collected and centrifuged for 5 min at 10,000g. Each pellet was resuspended in 1.0 ml of Trizol reagent (Invitrogen) and 100 $\mu$L of acid-washed glass beads were added (Sigma-Aldrich). Tubes were vortexed for 15 min at 4°C for cell lysis to occur. The tubes were then centrifuged for 10 min at 12,000g at 4°C and the supernatant was collected in a fresh 2-mL tube. 200 $\mu$L chloroform was then added and tubes were shaken by hand for 10 seconds. The tubes were again centrifuged for 10 min at 12,000g at 4°C. 400 $\mu$L of the upper aqueous phase was transferred to a new tube, and an equal volume of phenol-chloroform-isoamyl alcohol (pH 4.7) (Ambion, Austin, TX) was added. Tubes were again shaken by hand for 10 seconds and centrifuged for 10 min at 12,000g at 4°C. 250 $\mu$L of the upper phase was transferred to a new tube with an equal volume of cold ethanol and 1/10th volume sodium acetate (pH 5.2). Tubes were chilled at -20°C for thirty minutes to promote precipitation. Tubes were then centrifuged for 5 min at 12,000g, washed twice with 70% ethanol, dried in a 60°C oven and finally resuspended in RNAse free water. RNA quantity was analyzed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and samples were stored in -80°C freezer. qRT-PCR analyses were carried out using iScript One-step RT-PCR Kit with SYBR Green (Bio-Rad, Hercules, CA) using the Bio-Rad iCycler iQ Real-Time PCR Detection System. Fluorescence results were analyzed using Real-time PCR Miner and relative quantification and statistical analysis

was determined with REST 2009 (Qiagen) using actin as the reference gene and MTYL038 as the reference strain (Zhao and Fernald, 2005). Samples were analyzed in quadruplicate.

### 8.2.4 Xylose transport assay

Cells were preinoculated in a tube containing YPD media overnight at 28 $^{o}$C. These cells were then inoculated into a shake flask containing 40 mL of YPD media and grown overnight at 28 $^{o}$C at 200 rpm. The following day, when the culture has reached a high OD, the aliquots of culture were harvested and centrifuged. The media was aspirated and the pellet was washed with sterile distilled water by resuspension, centrifugation, and reaspiration. Finally the cell pellet was resuspended in minimal media containing 6.7 g/L YNB and 20 g/L substrate. The suspension was then transferred to a new shake flask and grown at 28 $^{o}$C. HPLC and dry cell weight samples were taken at 0, 5, 10, 24, 48 and 72 hrs to obtain substrate consumption and cell growth profiles. From this data, kinetic paramters for growth and substrate consumption were estimated by least squares method. These values were used to calculate the specific uptake rate, $q$, for the substrate using the following equation: $q = \frac{\mu}{Y}$ where $\mu$ = specific growth rate; $Y$ = Yield (Biomass($X$)/Substrate($S$)).

### 8.2.5 Direct transesterification

For routine lipid quantification to determine relative lipid accumulation, a method for direct transesterification of cell biomass was used, adapted from the two-step base-then-acid-catalyzed direct transesterification method developed by Griffiths et al. (2010). A normalized quantity of cell culture was centrifuged and the media supernatant was removed. Samples were then stored in -20 $^{o}$C freezer or directly transesterified. The cell was then resuspended with the addition of 100 $\mu$L of hexane containing 10 mg/mL methyl tridecanoate internal standard. 500 $\mu$L 0.5 N sodium methoxide, prepared by the addition of sodium hydroxide to methanol, was then added to the sample. The sample was then vortexed for 1 hour at room temperature. Next 40 $\mu$L of sulfuric acid was carefully added to the sample, followed by the

addition of 500 $\mu$L of neat hexane. The sample was again vortexed at room temperature for another 30 minutes. 300 $\mu$L of the upper hexane layer was then transferred into a glass vial and run using the GC-FID, under standard operating conditions. Total lipid content was calculated as the sum of total fatty acid content for the five primary FAMEs identified.

## 8.3 Results and Discussion

### 8.3.1 Elucidating endogenous functionality of the xylose utilization pathway in *Y. lipolytica*

Within the literature, there are conflicting reports about the ability for *Y. lipolytica* to naturally consume xylose. In most reports, growth on xylose has not been observed (Pan et al., 2009; Ruiz-Herrera and Sentandreu, 2002). However, there are reports of *Y. lipolytica* positively growing on xylose: strain Po1g was found to consume xylose in a cane hydrolysate fermentation (Tsigie et al., 2011), and two strains of *Y. lipolytica* were grown on xylose to measure xylulose-5-phosphate phosphoketolase activity (Evans and Ratledge, 1984). Beyond these incidences, there is otherwise very little reported evidence of using *Y. lipolytica* for growth on xylose, despite the volume of research of using the organism to grow on other alternative and residual substrate sources (Scioli and Vollaro, 1997; Papanikolaou et al., 2002, 2003). Table 8.3 lists putative XR/XDH/XK genes within the genome of *Y. lipolytica* from a BLAST comparison to known functional pathway genes. While the amino acid identity is only 40-52%, the expect value indicates significant likelihood of similarity, and *Y. lipolytica* often manages only 40-60% amino acid identity with orthologous genes from *S. cerevisiae*, due to distal phylogeny. Nonetheless, the low homology calls into question the potential functional characteristics of these genes, which further adds to the controversy.

To test the ability for *Y. lipolytica* to utilize its endogenous putative XYL123 pathway in laboratory conditions, control strain MTYL038 was grown in minimal media on three different substrates: xylose, xylitol, arabitol. As seen in Figure 8.3.1A, these three substrates

Table 8.3: BLAST results for endogenous xylose utilization pathway in *Y. lipolytica*. Amino acid identity is indicated in comparison with the parent sequence (organism indicated in parentheses). Expect value is the statistical false-positive rate.

| Function | Accession Number | Identity | Expect Value |
|---|---|---|---|
| Xylose reductase (XR) | YALI0D07634p | 49% (*S. stipitis*) | 3e-80 |
| Xylitol dehydrogenase (XDH) | YALI0E12463p | 52% (*S. stipitis*) | 1e-96 |
| Xylulokinase (XK) | YALI0F10923p | 40% (*S. cerevisiae*) | 1e-96 |

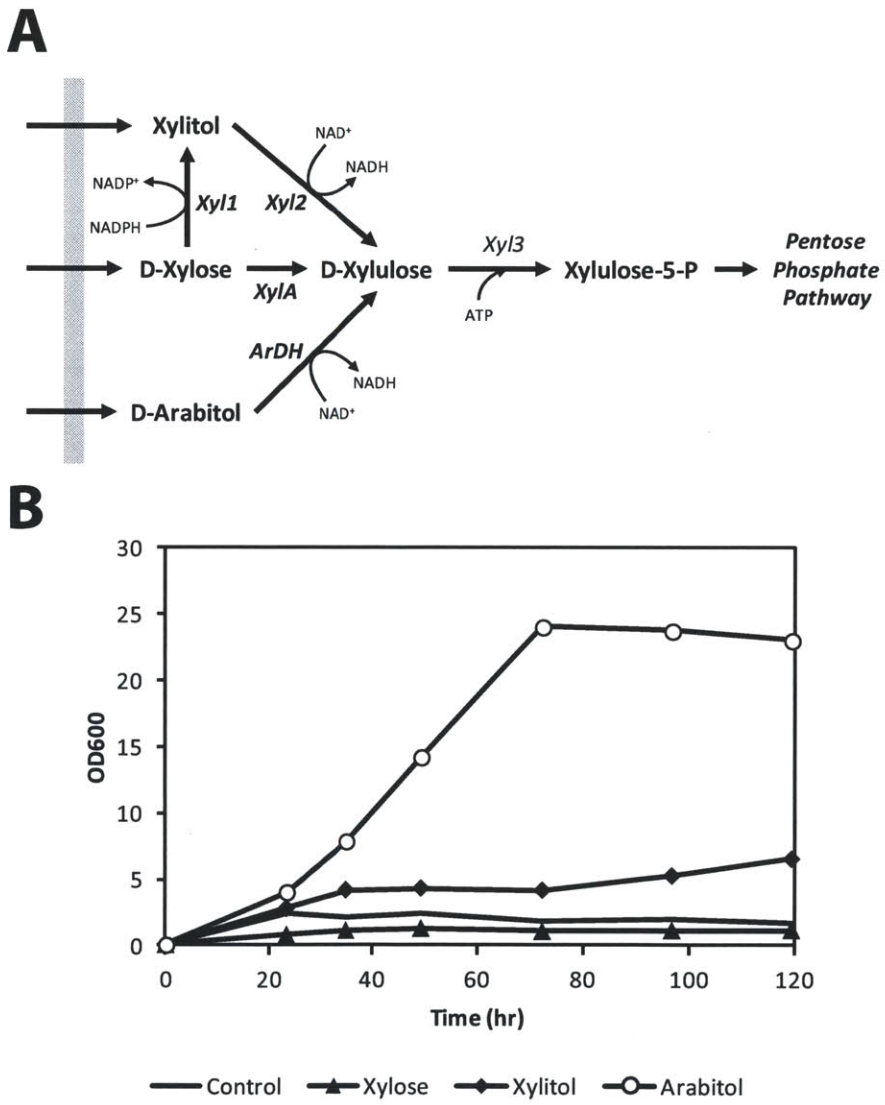


Figure 8.3.1: Diagnosing the functionality of endogenous xylose utilization genes. (A) Diagram of utilization pathways for xylose, xylitol, and D-arabitol. (B) Shake flask experiments with control strain MTYL038 grown on these substrates demonstrate growth on D-arabitol, poor growth on xylitol, and no growth on xylose.

can be used to diagnose the functionality of the three XYL123 genes. For example, growth on xylitol will demonstrate that XYL2 and XYL3 are functional, while growth on arabitol demonstrates that XYL3 is functional. Figure 8.3.1B depicts the growth curves of MTYL038 on the various substrates, with a shake flask with no carbon substrate as the control. While it was found that the strain did not grow on xylose, it was found to grow weakly on xylitol and quite robustly on arabitol. This suggests that while XYL1, and most likely XYL2, are not naturally expressed or functional in *Y. lipolytica* in the presence of their respective substrates, XYL3 is expressed and the organism can grow utilizing this pathway as its primary catabolic pathway.

### 8.3.2 Expression of XYL12 enables growth on xylose

With the knowledge that the endogenous xylulokinase is functional in *Y. lipolytica*, the remaining elements of the xylose utilization pathway were integrated to enable growth on xylose. The XYL1 and XYL2 genes from *S. stipitis* were amplified from plasmid p426-XYL123 and transferred into *Y. lipolytica* expression cassettes. XYL1 was cloned under the control of the stronger TEFin promoter, while the XYL2 gene was cloned under the control of hp4d. The XYL2 expression cassette was inserted into the XYL1 plasmid, creating plasmid pMT081, expressing both XYL1 and XYL2. Transformation of this plasmid into background strain Po1g yielded the strain MTYL081.

Numerous experiments working with *S. cerevisiae* and the xylose utilization pathway have discovered that it is often necessary to include periods of adaptation - where serial dilution in xylose media is performed - for development of stable xylose utilization (Jeffries, 2006; Kuyper et al., 2004; Tomás-Pejó et al., 2010). This was similarly found to be the case in *Y. lipolytica* - the verified transformant MTYL081 initially did not grow on xylose. It was grown in minimal xylose media in a shake flask for 10 days before reinoculating in fresh media. This serial dilution was repeated until there was an observed increase in maximum OD to above 15. Figure 8.3.2A shows the growth curve on the third serial dilution compared to the original

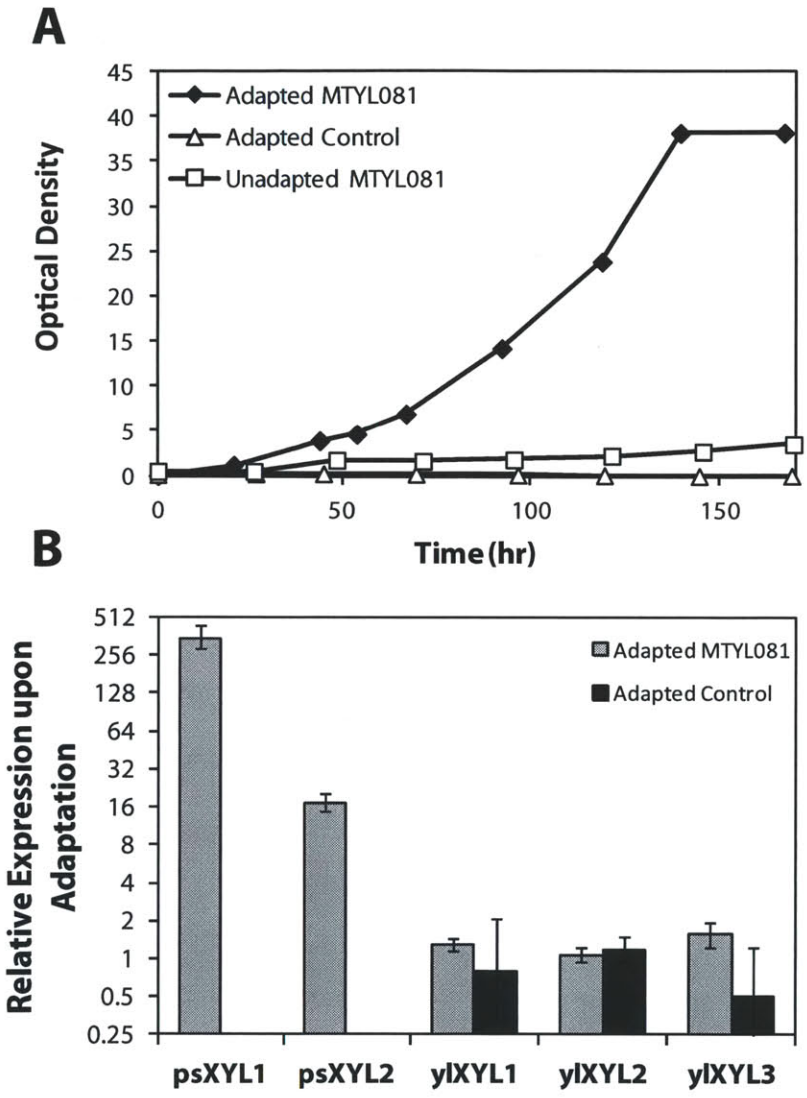


Figure 8.3.2: (A) Growth of adapted *Y. lipolytica* strain MTYL081 on xylose as sole carbon source in minimal media shake flask, compared to unadapted MTYL081 and control strain MTYL038 that underwent the adaptation protocol. (B) Transcriptional comparison of the xylose utilization pathway of an adapted *Y. lipolytica* strain and an unadapted strain. psXYL1 and psXYL2 are heterologously expressed from *S. stipitis*, while ylXYL1, ylXYL2, ylXYL3 are the endogenous putative xylose utilization pathway.

unadapted strain and a control strain that underwent serial dilution in xylose media. Lack of growth from the latter two strains shows that adaptation is necessary for xylose utilization and adaptation does not occur in strains lacking the heterologous XYL12 genes. Adapted growth was found to be steady and roughly exponential, with the maximum OD of 38 being reached after 130 hours. The doubling time is roughly 25 hrs, which is significantly lower than rates typically observed on glucose but comparable to that on arabitol (see Figure 8.3.1B).

To explore the underlying adaptations that improved the xylose-utilizing phenotype, RT-PCR was performed comparing the expression of heterologously expressed XYL12 and endogenous XYL123 genes in the adapted and unadapted strains. Figure 8.3.2B shows the relative change in transcription level of the genes after adaptation. The heterologously expressed XYL1 was overexpressed 300-fold compared to the unadapted strain, while XYL2 was upregulated 17-fold. Within the adapted strain, XYL1 was expressed 6-fold greater than XYL2, which is in agreement with the expression expected from the promoters used. Endogenous XYL123 was not significantly upregulated both in adapted MTYL081 and the control strain that underwent serial dilution, indicating that the observed adaptation to xylose was not an activation of the putative native xylose pathway. The strong upregulation of XYL1 and XYL2 has been similarly observed in metabolic engineering of *S. cerevisiae*, as the utilization pathway, being both heterologously expressed and potentially the rate-limiting step, requires strong overexpression for sufficient growth (Karhumaa et al., 2005, 2007). This seems to likewise be the case in *Y. lipolytica*, as the two XYL12 steps achieve very strong overexpression and yet still only achieve a relatively low growth rate. However, it may also be that with the adapted XYL12 expression, new rate-limiting steps appear to hinder specific growth on xylose, such as PPP activity or pentose transport (Karhumaa et al., 2005).

The normal combined activity of XYL1 and XYL2 consumes one NADPH and generates one NADH. Without suitable means to regenerate NADPH from NADH, this can lead to

cofactor imbalances and has been seen as a significant challenge in metabolic engineering of *S. cerevisiae* (Matsushika et al., 2009). However, with a potential cofactor imbalance, one would expect early cessation of growth and large accumulation of xylitol due to complete depletion on NADPH. In our shake flask cultures we observed only $< 0.5$ g/L xylitol formation after consumption of 32 g/L of xylose, while the maximum OD was very higher compared to what is typically observed in shake flasks, suggesting that cofactor balance may not be an issue in this situation. While this does not remove the possibility of rate-limiting steps in the exchange of NADPH to NADH, thus slowing but not stopping growth, in the presence of oxygen, mitochondrial function actively controls and maintains the NADPH/NADH equilibrium and exchange fluxes (Singh and Mishra, 1995).

### 8.3.3 Modeling growth and uptake on glucose and xylose

To understand the uptake characteristics of strain MTYL081 growing on xylose, an experiment was conducted to model the growth behavior and specific uptake rate, comparing growth on xylose and glucose. To minimize lag phase commonly observed on xylose growth, cells were initially grown on YPD medium (containing glucose) and were transferred at higher densities (OD 2) to minimal substrate medium after washing with distilled water.

After transfer to minimal medium containing glucose or xylose, both cultures resumed growth with no discernable lag phase. The growth profile and substrate conusmption are shown in Figure 8.3.3A and 8.3.3B, respectively. Growth on glucose was shown to completely consume all 20 g/L after 72 hours, while 3 g/L xylose remained at the same point in time in the xylose culture. While roughly the same biomass yield and production was achieved, growth on glucose and xylose differed dramatically in their growth kinetics. Growth on glucose followed a traditional substrate-limited growth with monod-like kinetics, i.e. initial exponential growth followed by population saturation. Growth on xylose, however, best followed a polynomial $x^2$ model. To investigate the underlying factors contributing to this difference, the specific uptake rate was calculated for the two systems. As shown in Figure

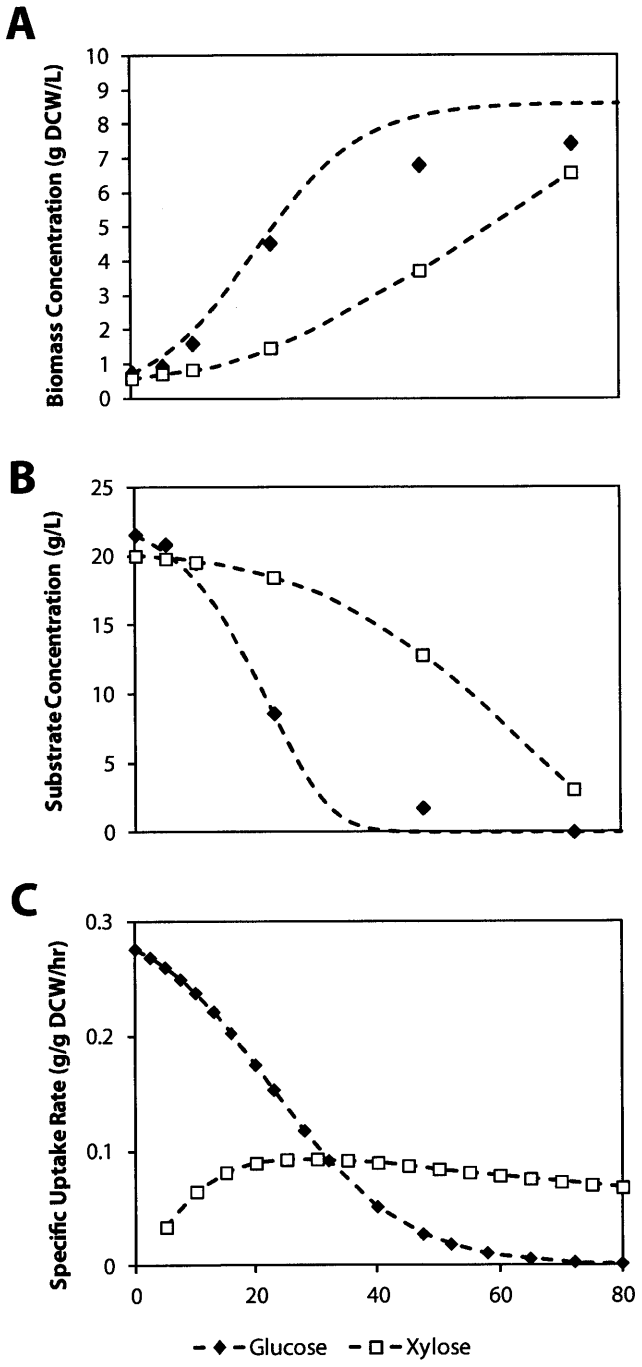


Figure 8.3.3: Specific Uptake Rate of XYL12 on glucose (black diamonds) or xylose (white squares). (A) The biomass production of strain MTYL081 in minimal media containing glucose or xylose. (B) The substrate concentration of the two cultures. (C) The calculated specific uptake rate of glucose and xylose during these fermentations.

8.3.3C, the substrate limited-growth on glucose has a maximum specific uptake rate at the beginning of the culture at 0.275 g glucose/g DCW/hr, but as substrate becomes depleted, specific consumption decreases alongside growth. On xylose, consumption stays roughly constant throughout the entire fermentation, with a $q_{max}$ of 0.093 g xylose/g DCW/hr, one third the maximum uptake rate on glucose. While the difference in substrate-sensitive and constant uptake rate have dramatic differences on the growth profile of the organism, the three-fold difference in maximum uptake rate is primarily responsible for low growth rate of the culture. The source and cause of low uptake is unclear, and the identification and alleviation of these bottlenecks has been a common and recurring challenge in engineering xylose utilization pathways Matsushika et al. (2009). The abrupt maximum reached in Figure 8.3.2A seems to similarly demonstrate the substrate-insensitive, constitutive uptake rate observed.

### 8.3.4 Cofermentation of two substrates for improved productivity

While metabolic engineering allowed growth on xylose in *Y. lipolytica*, growth was dramatically slower than on glucose. Possible factors contributing to the limited growth and productivity are the lack of dedicated pentose transporters, low PPP flux, and inability for the cell to identify xylose as a fermentable sugar (Jin et al., 2004; Jeffries, 2006; Matsushika et al., 2009). To improve productivities with the limited specific growth on xylose, experiments were performed using two-substrate cofermentations. Cellulosic materials typically consist of a blend of both hexose and pentose sugars, and rarely consist of pure pentose (Lee et al., 2007). Furtheremore, substrates like glycerol are a byproduct of biodiesel production, and may be recycled back into the process. First it was necessary to characterize and determine which cofermentation combinations are ideal for lipid production. Xylose was combined with a helper substrate - glucose, glycerol, or arabitol - and grown in shake flasks to determine growth characteristics and observe catabolite repression effects in the cofermentation system. Catabolite repression is the preferential uptake of one substrate through the repression of the

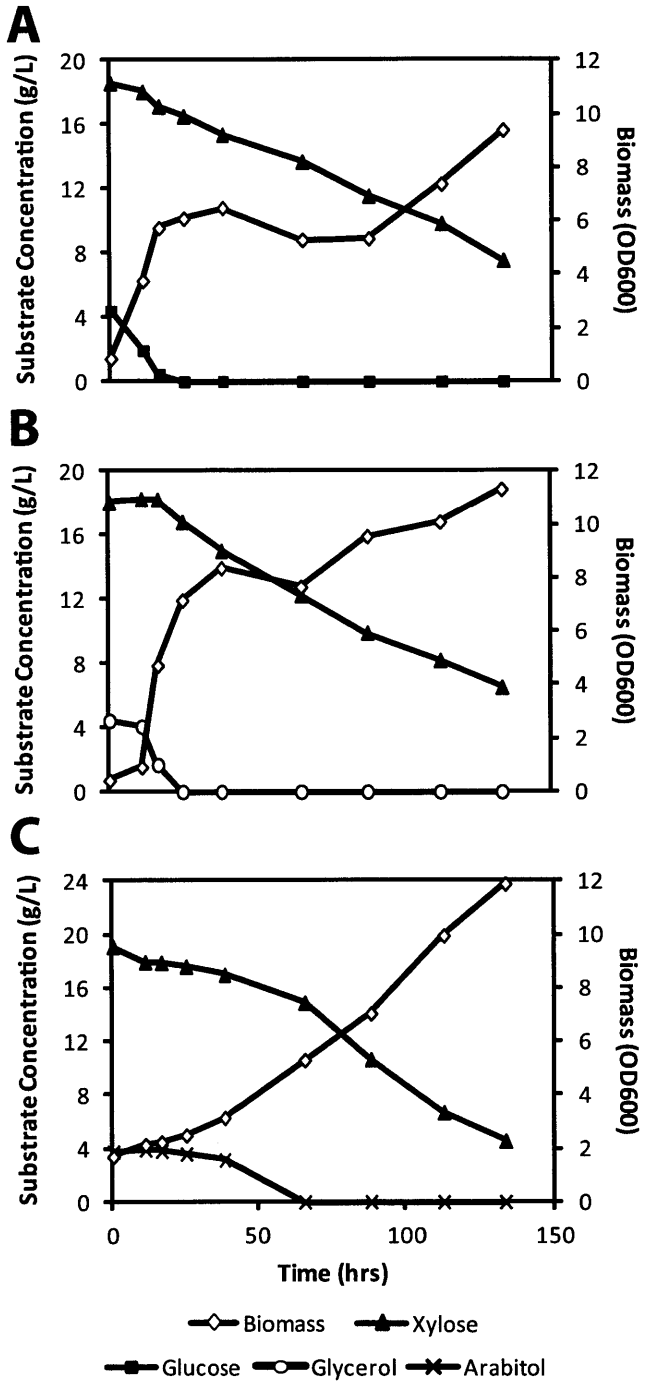


Figure 8.3.4: Cofermentation of xylose with glucose (A) , glycerol (B), or D-arabitol (C). Cultures were grown on 20 g/L xylose and 4 g/L of the secondary substrate.

utilization pathway of secondary substrates, and can be seen in a wide range of cofermentations in *Y. lipolytica* (Morgunov and Kamzolova, 2011). The strain MTYL085 was used, which contains the XYL12 pathway as well as DGA overexpression. DGA overexpression is capable of improving lipid accumulation by 3-fold and was found to be the most significant contributor to engineered lipid overproduction (See Chapter 6). By combining both the xylose utilization pathway and elements for lipid overproduction, we may be able to direct flux from xylose towards lipids for a cellulosic biodiesel platform.

Figure 8.3.4 depicts the growth characteristics and depletion of both substrates for the three cofermentation combinations. For glycerol (Figure 8.3.4B), diauxic shift is clearly observed, with glycerol being consumed rapidly before any xylose is depleted. For glucose (Figure 8.3.4A), diauxic shift was less observable, as it is possible that at very low concentrations of glucose, catabolite repression is weak (Morgunov and Kamzolova, 2011). At higher glucose concentrations, diauxic shift was clearly observable (data not shown). While all three cultures began with 4 g/L of the helper substrate, glycerol was converted into the most biomass after it was completely depleted, achieving an OD of 8 within 24 hrs. Glycerol has been known to be a highly preferred substrate for *Y. lipolytica*, and unlike *S. cerevisiae*, there is no loss in specific growth rate when growing on glycerol compared to glucose (Taccari et al., 2012). It is also Crabtree-negative, an effect that eschews the respiration-dependent nature of glycerol metabolism found in *S. cerevisiae* (De Deken, 1966). As a result, MTYL085 is able to consume slightly more xylose by the end of the culture. The evidence of diauxic shift also indicates that while the xylose uptake rate may be constant when grown solely on xylose, other factors must be at play in repressing the utilization, most conspicuously pentose transport. There is a growing body of evidence that pentose transport is a key rate-limiting step in xylose utilization and may also be a strong contributing factor towards diauxic shift (Young et al., 2012).

The cofermentation of xylose and arabitol exhibits a much different response (Figure 8.3.4C). Since arabitol shares the same catabolic route for all but the initial pathway, it is

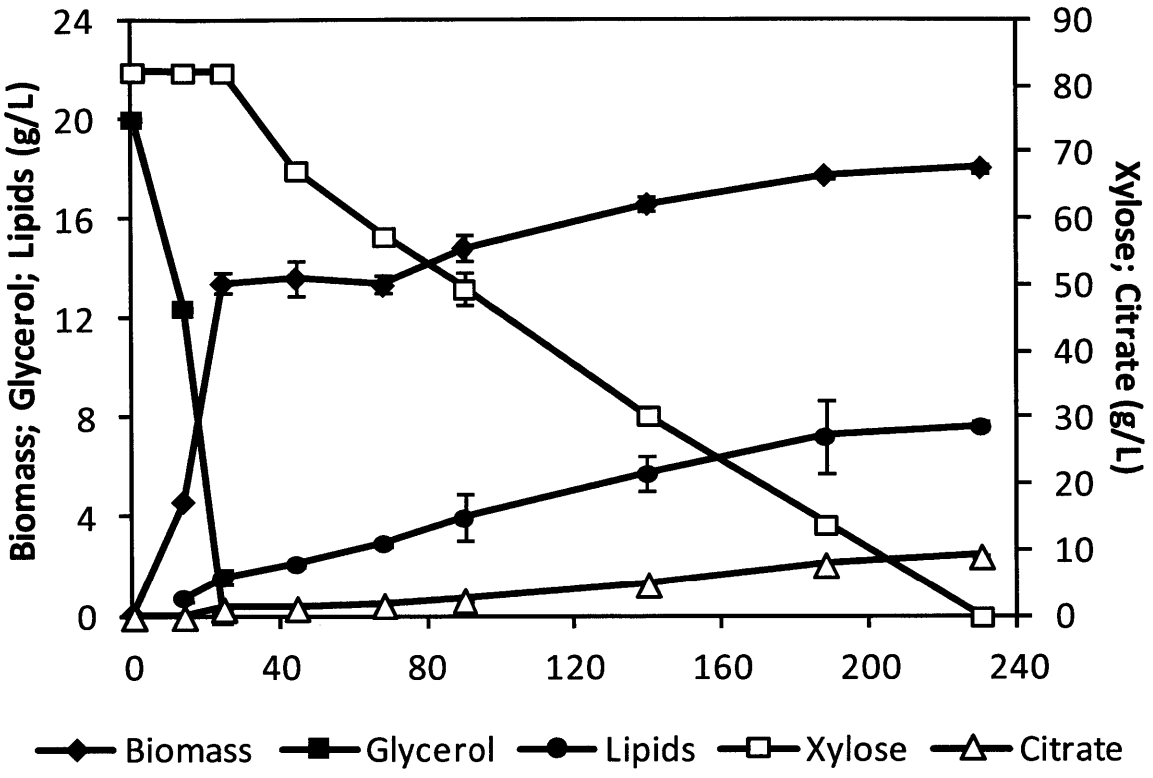


Figure 8.3.5: 2-L bioreactor fermentation of strain MTYL081 on glycerol and xylose.

likely the arabitol response will be most similar to the xylose growth phenotype. Indeed, arabitol cofermentation exhibited the same parabolic growth profile observed in Figure 8.3.3. Furthermore, xylose depletion begins well before arabitol is consumed, exhibiting simultaneous utilization of both substrates. The smooth growth profile in this case is in contrast to the two-phase growth seen in glucose or glycerol - a product of diauxic growth. Nonetheless the overall growth rate and productivity is significantly lower than glucose or glycerol. Additionally, arabitol is not a common substrate in cellulosic material and would thus be a prohibitive cost to supplement as a feedstock.

### 8.3.5 Lipid production in xylose and glycerol cofermentation

Because glycerol showed the greatest promise for increased productivity, a scale-up cofermentation was performed using glycerol and xylose as substrates. A 2-L bioreactor was

initially charged with 20 g/L glycerol and 80 g/L xylose. The C/N ratio of the reactor was adjusted to be 100, which results nitrogen-limited conditions favorable for lipid accumulation. The results of the fermentation are found in Figure 8.3.5. Over the course of 230 hrs, all the carbon substrate was consumed, with glycerol being depleted within the first 24 hrs. Diauxic shift can clearly be observed, as no xylose is consumed until after all the glycerol has been depleted. The 20 g/L of glycerol was able to generate 13 g/L of biomass. Lipid accumulation steadily occurred between 70 and 230 hours, with a majority of the biomass generated on xylose being accounted as lipids. The culture finally achieved a biomass concentration of 18 g/L with 7.64 g/L lipids, or 42% of total biomass. While this is lower than the 62% achieved on glucose using strain MTYL065 (expressing both ACC and DGA), the single expression of DGA was able to convert xylose into lipids at quantities similar to other *Y. lipolytica* fermentations (Beopoulos et al., 2009; Papanikolaou and Aggelis, 2002). The fatty acid profile was also found to be similar to grown on glucose with about 50% of total fatty acids in the form of oleate (data not shown). The yield of lipid production, however, was very low. Of the 80 g/L of xylose consumed, only 6.08 g/L of lipids was generated, for a yield of 0.074 g lipids/g xylose. This is only 21.7% of the theoretical yield, and was very low even compared to biomass yield in the shake flask experiments of Figure 8.3.3. This low yield is presumably due to overrespiration of the carbon substrate. It may be possible that the high aeration of the bioreactor fermentation could have adversely affected the respiratory response on xylose not observed in shake flask, thus resulting in very low yields of biomass and lipids. Furthermore, 9.13 g/L citrate was also generated, which actually accounts for a significant yield from the 100 g/L of carbon substrate initially charged. It is possible that the C/N ratio was too high, as extreme C/N ratios in *Y. lipolytica* fermentations can tend to produce citrate instead of lipids, likely due to limited ability to generate sufficient ATP for fatty acid synthesis (Beopoulos et al., 2009). Despite these low yields, the vast majority (80%) of the lipids were produced during the xylose-only phase, indicating successful con-

version of xylose-to-lipids using *Y. lipolytica*, a first step in developing a cellulosic biodiesel platform.

### 8.3.6 Transcriptional expression affected by secondary substrate

To further investigate the response of *Y. lipolytica* during cofermentations with xylose and the overrespiration observed on glycerol-xylose, transcriptional analysis was performed on genes within the TCA cycle. Xylose consumption in *S. cerevisiae* elicits a non-fermentative response and general upregulation of the TCA cycle (Jin et al., 2004; Salusjarvi et al., 2006). This results in lower efficiencies in xylose utilization for ethanol production as downregulation of the TCA cycle is necessary to divert carbon flux towards ethanol fermentation, whether via anaerobic environmental conditions or activity of the Crabtree effect. In our cofermentation system, the response of *Y. lipolytica* when transitioning from the helper substrate to xylose was examined. An initial RNA extraction was performed during the cofermentation while still growing on glucose, glycerol or arabitol, and a second RNA extraction was performed after the helper substrate was depleted and the strain was exhibiting growth on xylose as sole carbon substrate. From this we can identify if a similar respiratory response is observed on xylose. Figure 8.3.6 depicts the fold-change in transcripts for pyruvate dehydrogenase (PDB1, Accession Number: XM_504448), Aconitase (ACO1, Accession Number: XM_502616), isocitrate lyase (ICL, Accession Number: XM_501923), and isocitrate dehydrogenase (IDH1, Accession Number: XM_503571). These genes represent key enzymatic steps for the utilization of TCA cycle intermediates: PDB1, entrance into the TCA cycle; ACO1, diverting citrate to the TCA cycle instead of the cytosol; ICL1, diverting isocitrate through the glyoxylate shunt; IDH1, committed step into oxidative respiration.

In all three cases, PDB1 is significantly upregulated, suggesting that there is a stronger driving force towards the TCA cycle in xylose than any other substrate. Aconitase overexpression was not observed in the glucose-to-xylose transition, but was dramatically increased 50-fold in the glycerol-to-xylose transition. This was mostly due to very low transcription

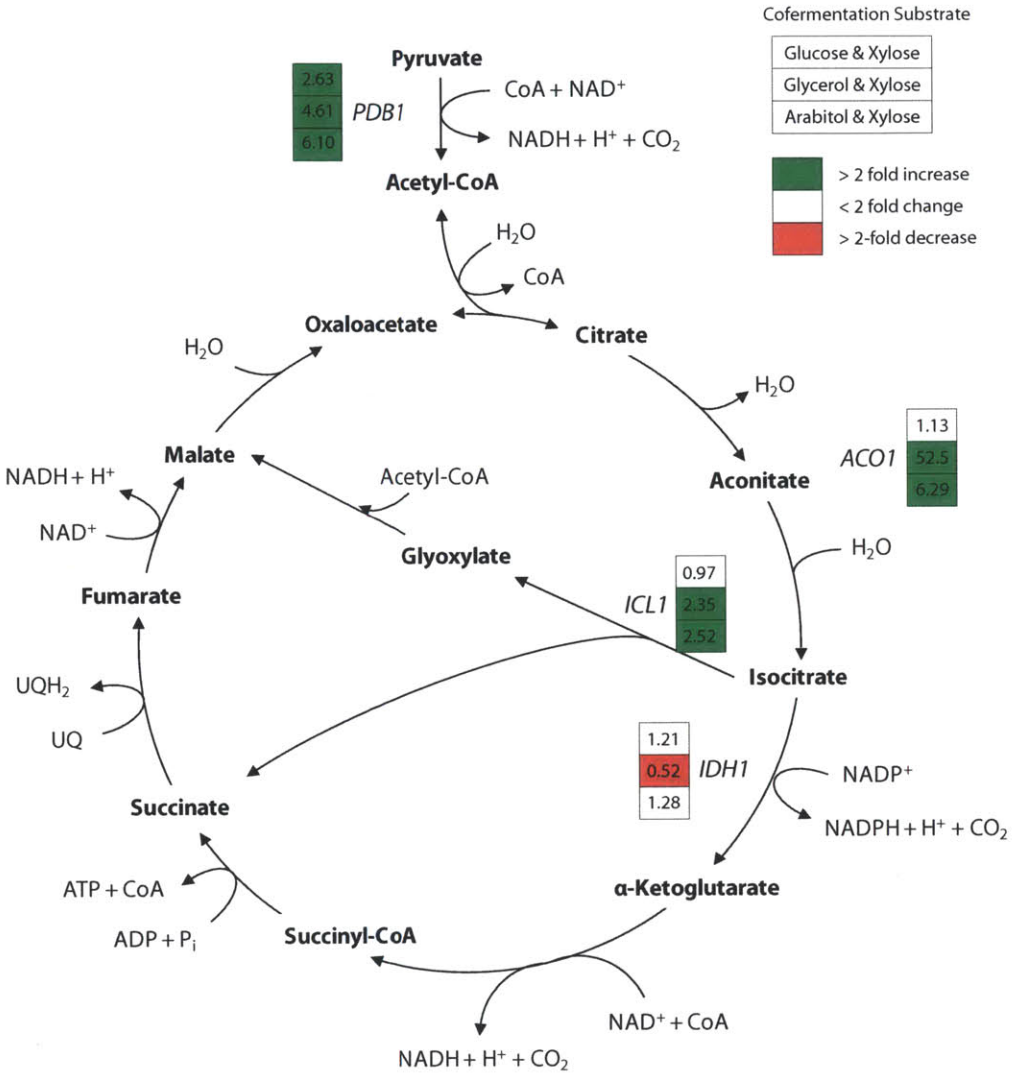


Figure 8.3.6: Comparison of mRNA levels of genes responsible for energy production during xylose cofermentation with a secondary substrate: glucose, glycerol, arabitol. The comparison is between two time points during the cofermentation: when primarily the secondary substrate is being consumed vs. when the secondary substrate is depleted and only xylose is being consumed. Transcript levels that did not change significantly are shown in white boxes. Transcript levels that increased more than two-fold after transitioning to xylose utilization are shown in green boxes. Transcript levels that decreased more than two-fold after transitioning to xylose are shown in red boxes. Numbers inside of each box indicate the ratio of transcripts during the xylose-only phase vs. secondary substrate phase. Numbers greater than 1.0 signify up-regulation of the gene when transitioning from secondary substrate to xylose, while numbers less than 1.0 signify downregulation.

levels observed of ACO1 on glycerol rather than extraordinarily high expression of ACO1 on xylose. ACO1 was upregulated in the transition from arabitol to xylose as well. For ICL1, significant increase in expression was observed during the glycerol-to-xylose transition and the arabitol-to-xylose transition, but not on glucose. In most organisms, ICL1 is normally not expressed due to strong catabolite repression; however, *Y. lipolytica* seems to exhibit constitutive expression of the pathway (Flores and Gancedo, 2005). Indeed, the magnitude of changes in expression of ICL1 suggests significant expression prior to the transition. Finally, IDH1 expression is not significantly changed in glucose and arabitol, but is actually downregulated on glycerol, indicating that respiration is much more strongly upregulated on glycerol than xylose.

The upregulation of PDB1 and ACO1 in the glycerol fermentation demonstrate an elevated respiratory response when transitioning from glycerol to xylose utilization. While IDH1 is downregulated, the upstream regulation may be enough to result in the overrespiration observed in the bioreactor. It is unclear why ACO1 is downregulated so dramatically when growing on glycerol, but any previous regulation on this enzyme must surely be alleviated. On the other hand, glucose-xylose cofermentation resulted in few significant changes in transcription. This may indicate that glucose-xylose cofermentation may yield better results at larger scales despite the stronger preference for glycerol by *Y. lipolytica*.

## 8.4 Conclusion

Pentose utilization represents a pressing need in the development of sustainable biofuel production, as the push and advantages for cellulosic feedstocks begin to outweigh the technical challenges. The oleaginous yeast *Y. lipolytica* is an example of a robust platform for the production of yeast oil that can be converted into biodiesel. Through metabolic engineering, the robust lipid production capabilities established in *Y. lipolytica* can be expanded to include xylose utilization, enabling further opportunities for microbial cellulosic biodiesel produc-

tion. By testing native growth on a variety of substrates we showed that the endogenous XYL3 is functional in minimal medium, while the putative XYL12 genes are not. Through heterologous expression of XYL1 and XYL2 genes from *S. stipitis* we enabled xylose utilization in *Y. lipolytica* after an adaptation period. Through cofermentation we were able to eliminate lag phases and increase growth and productivity on xylose, ultimately achieving 42% lipid accumulation in a strain that is metabolically engineered in both xylose utilization and lipid accumulation pathways. By observing the TCA cycle response, we also observed variation between cofermentation substrates, suggesting a transcriptional regulatory basis for overrespiration. By leveraging the knowledge base developed from the study of xylose utilization in *S. cerevisiae*, these results establish a framework for studying and engineering the oleaginous yeast *Y. lipolytica* for xylose utilization and the production of cellulosic biodiesel.

# References


Barth, G., and Gaillardin, C. (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica. FEMS Microbiol. Rev. 19*, 219–237.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J. L., Molina-Jouve, C., and Nicaud, J. M. (2009) *Yarrowia lipolytica* as a model for bio-oil production. *Progress in Lipid Research 48*, 375–387.

Blank, L. M., Lehmbeck, F., and Sauer, U. (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res. 5*, 545–558.

Chen, D. C., Beckerich, J. M., and Gaillardin, C. (1997) One-step transformation of the dimorphic yeast *Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 48*, 232–235.

De Deken, R. (1966) The Crabtree effect and its relation to the petite mutation. *Journal of general microbiology 44*, 157.

Evans, C. T., and Ratledge, C. (1984) Induction of xylulose-5-phosphate phosphoketolase in a variety of yeasts grown ond-xylose: the key to efficient xylose metabolism. *Arch. Microbiol. 139*, 48–52.

Flores, C.-L., and Gancedo, C. (2005) *Yarrowia lipolytica* Mutants Devoid of Pyruvate Carboxylase Activity Show an Unusual Growth Phenotype. *Eukaryotic Cell 4*, 356–364.

Griffiths, M. J., van Hille, R. P., and Harrison, S. T. L. (2010) Selection of direct transesterification as the preferred method for assay of fatty acid content of microalgae. *Lipids 45*, 1053–1060.

Jeffries, T. W. (2006) Engineering yeasts for xylose metabolism. *Curr. Opin. Biotechnol. 17*, 320–326.

Jin, Y.-S., Laplaza, J. M., and Jeffries, T. W. (2004) *Saccharomyces cerevisiae* Engineered for Xylose Metabolism Exhibits a Respiratory Response. *Appl. Environ. Microbiol. 70*, 6816–6825.

Karhumaa, K., Hahn-Hägerdal, B., and Gorwa-Grauslund, M. (2005) Investigation of limiting metabolic steps in the utilization of xylose by recombinant *Saccharomyces cerevisiae* using metabolic engineering. *Yeast 22*, 359–368.

Karhumaa, K., Sanchez, R., Hahn-Hagerdal, B., and Gorwa-Grauslund, M.-F. (2007) Comparison of the xylose reductase-xylitol dehydrogenase and the xylose isomerase pathways for xylose fermentation by recombinant *Saccharomyces cerevisiae. Microbial Cell Factories 6*, 5.

Kuyper, M., Winkler, A. A., van Dijken, J. P., and Pronk, J. T. (2004) Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: a proof of principle. *FEMS Yeast Res. 4*, 655–664.

Lee, D., Owens, V. N., Boe, A., and Jeranyama, P. *Composition of herbaceous biomass feedstocks*; South Dakota State University, 2007; Vol. 16.

Matsushika, A., Inoue, H., Kodaki, T., and Sawayama, S. (2009) Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Appl. Microbiol. Biotechnol. 84*, 37–53.

Morgunov, I. G., and Kamzolova, S. V. Yarrowia Lipolytica Yeast Possesses An Atypical Catabolite Repression. 2011.

Pan, L. X., Yang, D. F., Li, S., Wei, L., Chen, G. G., and Liang, Z. Q. (2009) Isolation of the Oleaginous Yeasts from the Soil and Studies of Their Lipid-Producing Capacities. *Food Technology and Biotechnology 47*, 215–220.

Papanikolaou, S., and Aggelis, G. (2002) Lipid production by *Yarrowia lipolytica* growing on industrial glycerol in a single-stage continuous culture. *Bioresour. Technol. 82*, 43–49.

Papanikolaou, S., Chevalot, I., Komaitis, M., Marc, I., and Aggelis, G. (2002) Single cell oil production by *Yarrowia lipolytica* growing on an industrial derivative of animal fat in batch cultures. *Appl. Microbiol. Biotechnol. 58*, 308–312.

Papanikolaou, S., Muniglia, L., Chevalot, I., Aggelis, G., and Marc, I. (2003) Accumulation of a cocoa-butter-like lipid by *Yarrowia lipolytica* cultivated on agro-industrial residues. *Curr. Microbiol. 46*, 124–130.

Perlack, R. D. *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasability of a Billion-Ton Annual Supply*; 2005.

Ratledge, C. In *Single Cell Oil*; Moreton, R., Ed.; Longman Scientific & Technical, 1988; Chapter Biochemistry, stoichiometry, substrate and economics, pp 33–70.

Ruiz-Herrera, J., and Sentandreu, R. (2002) Different effectors of dimorphism in *Yarrowia lipolytica. Arch Microbiol 178*, 477–483.

Salusjarvi, L., Pitkanen, J. P., Aristidou, A., Ruohonen, L., and Penttila, M. (2006) Transcription analysis of recombinant *Saccharomyces cerevisiae* reveals novel responses to xylose. *Appl. Biochem. Biotechnol. 128*, 237–261.

Sambrook, J., and Russell, D. W. *Molecular cloning: a laboratory manual*; CSHL press, 2001; Vol. 2.

Scioli, C., and Vollaro, L. (1997) The use of *Yarrowia lipolytica* to reduce pollution in olive mill wastewaters. *Water Res. 31*, 2520–2524.

Singh, A., and Mishra, P. *Microbial Pentose Utilization: Current Applications in Biotechnology*; Elsevier Science, 1995; Vol. 33.

Taccari, M., Canonico, L., Comitini, F., Mannazzu, I., and Ciani, M. (2012) Screening of yeasts for growth on crude glycerol and optimization of biomass production. *Bioresource Technology 1*, 1.

Tomás-Pejó, E., Ballesteros, M., Oliva, J., and Olsson, L. (2010) Adaptation of the xylose fermenting yeast Saccharomyces cerevisiae F12 for improving ethanol production in different fed-batch SSF processes. *Journal of Industrial Microbiology & Biotechnology 37*, 1211–1220.

Tsigie, Y. A., Wang, C.-Y., Truong, C.-T., and Ju, Y.-H. (2011) Lipid production from *Yarrowia lipolytica* Po1g grown in sugarcane bagasse hydrolysate. *Bioresour. Technol. 102*, 9216–9222.

Walfridsson, M., Hallborn, J., Penttilä, M., Keränen, S., and Hahn-Hägerdal, B. (1995) Xylose-metabolizing Saccharomyces cerevisiae strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase. *Applied and environmental microbiology 61*, 4184–4190.

Young, E. M., Comer, A. D., Huang, H., and Alper, H. S. (2012) A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae. Metab. Eng. 1*, 1.

Zhao, S., and Fernald, R. D. (2005) Comprehensive Algorithm for Quantitative Real-Time Polymerase Chain Reaction. *J. Comput. Biol. 12*, 1047–1064.

Zhou, H. Metabolic engineering of yeast for xylose uptake and fermentation. Ph.D. thesis, Massachusetts Institute of Technology, 2011.

# Chapter 9

# Exploring Acetate Utilization in *Yarrowia lipolytica*

## 9.1 Introduction

As the industry for renewable energy matures, foreseeable challenges still exist for the various energy technologies. Wind and solar energy technologies appear to be the most viable renewable energy technologies being developed (Jacobson, 2009). While electricity generated from these sources will develop to become more efficient and plentiful, the need for dense, liquid transportation fuels will still remain, for example in maritime transport and aviation. Current technologies for sustainable biofuel production require significant amounts of land for crop production. These requirements place extra demand on arable land and can increase competition for food production. Since the cost of feedstock can constitute a significant portion of the operating cost for production (up to 60%), the margin for economic and sustainable production can be highly sensitive to market pricing Stephanopoulos (2007).

In light of this, a promising research goal is the development of processes that can efficiently convert electricity into hydrocarbon transportation fuels by non-photosynthetic pathways. This completely eliminates the scaling and land requirements associated with biofuel production. When produced by microorganisms, these electrofuels offer the opportunity for scalable production of liquid fuels that can be tailored in their composition and characteristics specifically for the transportation industry (Biello, 2011). In nature, there are a number of microorganisms that have been discovered that can grow and proliferate from carbon dioxide and electricity as sole carbon and energy sources. For many of these organisms, the conversion of carbon dioxide and electricity produces acetate as a byproduct. These acetogens thus represent promising targets for metabolic engineering for the production of metabolic products and biofuels at potentially high yields.

These organisms, however, are poorly understood and tools for engineering are relatively nascent. Furthermore, the production of some biofuels, such as biodiesel, operate more productively under aerobic conditions, while most acetogens are strictly anaerobes. An approach to decouple these process limitations is through the use of a two-step fermentation system using two different organisms: one anaerobically converting electricity and carbon

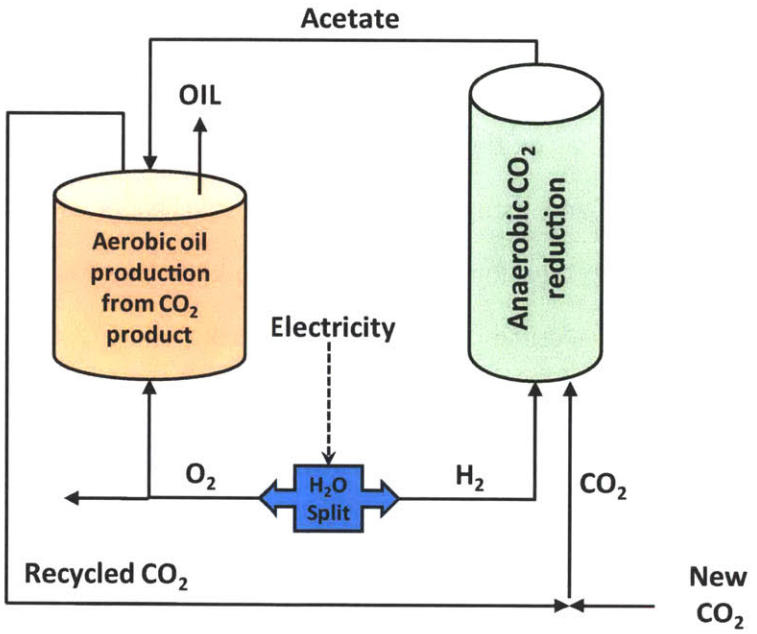


Figure 9.1.1: Process flow diagram for two bioreactor conversion of carbon dioxide into lipids. Carbon dioxide is fed into an anaerobic fermentation reactor for conversion to acetic acid using homoacetogenic organism. The acetic acid or acetate is fed into a second aerobic bioreactor as carbon substrate for lipid production using an oleaginous organism. Electrolysis is used to decompose water to form both hydrogen and oxygen, which are supplied to the anaerobic and aerobic reactors, respectively.

dioxide to produce high yields of acetate, and the other producing biofuel from the acetate generated from the first organism. The overall process thus is capable of converting electricity and carbon dioxide into biofuels, using acetate as the intermediate substrate. Figure 9.1.1 describes a hypothetical design for this process, where electrolysis of water produces both hydrogen and oxygen, which can be used in the first and second fermentations, respectively. Carbon dioxide and generated hydrogen enter the anaerobic reactor, and the acetogen utilizes these carbon and reducing substrates to generate acetate. This acetate is then fed into the second reactor along with oxygen where an oil-producing organism then converts this into lipids. The cells from the second bioreactor are then harvested for extraction of lipids and transesterified into biodiesel.

The oleaginous yeast *Y. lipolytica* is an ideal candidate for the lipid platform in the second bioreactor. As an oleaginous microorganism, it has numerous advantages with respect to

acetate to lipid conversion. *Y. lipolytica* has a well-studied and robust pathway for both lipid utilization and hydrophobic substrate degradation, able to rapidly grow off of these substrates. Since both $\beta$- and $\omega$- oxidation of fatty acids ultimately generate large pools of acetyl-coA, acetate is catabolized using much the same pathways as fatty acid utilization, only requiring one or two enzymes for coenzyme A (coA) activation. Indeed *Y. lipolytica* has been observed to grow on acetate (Barth and Gaillardin, 1997), and has even been used recently in the valorization of acetate into lipids (Fontanille et al., 2012).

A diagram of the entire lipid synthesis pathway, with acetate utilization, is shown in Figure 9.1.2. To generate energy, acetate is sent through the TCA cycle, where all the carbon is converted into carbon dioxide by the action of isocitrate dehydrogenase (IDH) and $\alpha$-ketoglutarate dehydrogenase. For regeneration of TCA cycle and glycolytic intermediates, acetate is sent through the glyoxylate shunt. Flux through the glyoxylate pathway is essential for utilizing the carbon from acetate(de la Peña Mattozzi et al., 2010), and in *Y. lipolytica* the controlling enzyme isocitrate lyase (ICL) is strongly upregulated in the presence of both fatty acids, acetate, or ethanol (Juretzek et al., 2001). After entering the cell, acetate is quickly converted into cytosolic acetyl-coA. As the primary precursor for fatty acid synthesis, it can be rapidly utilized in the lipid synthesis pathway.

To test the potential effectiveness of lipid production on acetate, we will use a metabolically engineered strain of *Y. lipolytica*, MTYL065, which has been engineered to overexpress acetyl-coA carboxylase (ACC) and diacylglycerol acyltransferase (DGA), and test its performance on lipid production from acetate as sole carbon source. These genetic modifications have been shown to dramatically increase the lipid production in *Y. lipolytica* on glucose (see Chapter 6 on page 93). As cytosolic acetyl-coA pools are heavily enriched during acetate consumption, the overexpression of ACC is ideally situated to direct flux towards lipid biosynthesis. Coupling this with the strong driving force for lipid sequestration established by DGA overexpression, strain MTYL065 has the potential to produce large quantities of lipids from acetate.

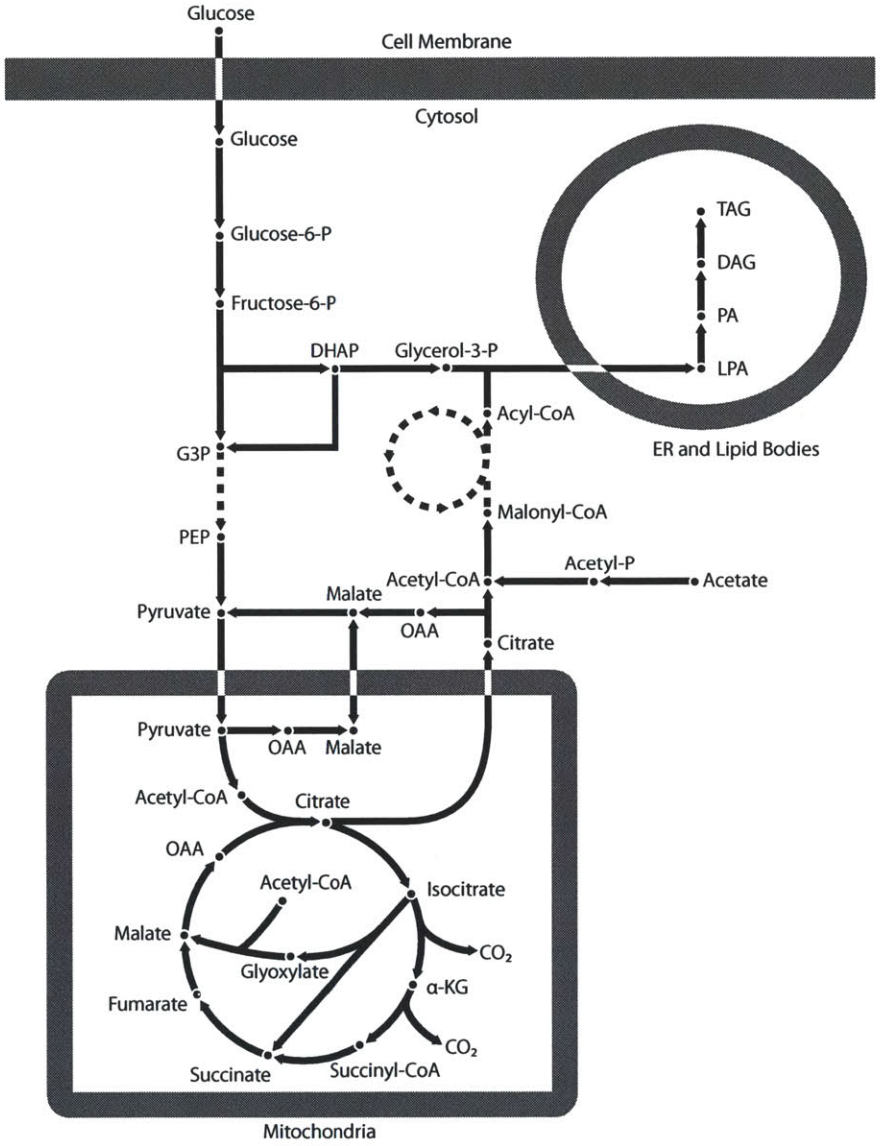


Figure 9.1.2: Overview of the principal metabolic pathways for lipid synthesis in *Y. lipolytica*. Acetate enters the cell and is converted to cytosolic acetyl-coA. Cytosolic acetyl-CoA is then converted into malonyl-CoA by acetyl-CoA carboxylase (ACC) as the first step of fatty acid synthesis. After fatty acid synthesis, triacylglycerol (TAG) synthesis follows the Kennedy pathway, which occurs in the endoplasmic reticulum (ER) and lipid bodies. Acyl-CoA is the precursor used for acylation to the glycerol-3-phosphate backbone to form lysophosphatidic acid (LPA), which is further acylated to form phosphatidic acid (PA). PA is then dephosphorylated to form diacylglycerol (DAG) and then a final acylation occurs by diacylglycerol acyltransferase (DGA) to produce TAG. The anapleurotic reactions responsible for converting acetate into TCA cycle intermediates is the glyoxylate shunt. OAA oxaloacetate, $\alpha$-KG alpha-ketoglutarate, PEP phosphoenolpyruvate, G3P Glyceraldehyde 3-phosphate, DHAP dihydroxyacetone phosphate.

Here we describe the use of a metabolically engineered strain of *Y. lipolytica* for the proof of concept of a two-step electrofuel process with acetate as the intermediate, converting acetate to lipids. We establish conditions for growth on acetate as the sole carbon source and evaluate its performance in a 2-L bioreactor. Finally we compare these results to performance of the same strain when grown on glucose. With these results, we establish *Y. lipolytica* as a strong and robust platform for lipid production from acetate, which can be used for research in electrofuel production.

## 9.2 Materials and Methods

### 9.2.1 Yeast strains, growth, and culture conditions

The *Y. lipolytica* strain used in this study was MTYL065, overexpressing ACC and DGA through the integration of plasmid pMT065, derived from base strain *Y. lipolytica* Polg (Yeastern Biotech). Information about the construction of this strain and plasmid can be found in Chapter 6 on page 93.

Shake flask experiments were carried out using the following medium: 1.7 g/L yeast nitrogen base (without amino acids), 1.5 g/L yeast extract, and 50 g/L glucose. From frozen stocks, precultures were inoculated into YNB medium (5 mL in Falcon tube, 200 rpm, 28°C, 24 hr). Overnight cultures were inoculated into 50 mL of media in 250 mL Erlenmeyer shake flask to an optical density (A600) of 0.05 and allowed to incubate for 100 hours (200 rpm, 28°C), after which biomass, sugar content, and lipid content were taken and analyzed.

Bioreactor scale fermentation was carried out in a 2-liter baffled stirred-tank bioreactor. The medium used contained 1.5 g/L yeast nitrogen base (without amino acids and ammonium sulfate), 2 g/L ammonium sulfate, 1 g/L yeast extract, and 90 g/L glucose. From a selective plate, an initial preculture was inoculated into YPD medium (40 mL in 250 mL Erlenmeyer flask, 200 rpm, 28°C, 24 hr). Exponentially growing cells from the overnight preculture were transferred into the bioreactor to an optical density (A600) of 0.1 in the

2-L reactor (2.5 vvm aeration, pH 6.8, 28°C, 250 rpm agitation). Time point samples were stored at -20°C for subsequent lipid analysis. Sugar organic acid content was determined by HPLC. Biomass was determined by determined gravimetrically from samples dried at 60°C for two nights.

### 9.2.2 Lipid extraction and quantification

Total lipids were extracted using the procedure by Folch et al (1957). A measured quantity of cell biomass (roughly 1 mg) was suspended in 1 mL of chloroform:methanol (2:1) solution and vortexed for 1 hour. After centrifugation, 500 $\mu$L was transferred to 125 $\mu$L saline solution. The upper aqueous layer was removed and the bottom layer was evaporated and resuspend in 100 $\mu$L hexane. Samples were then stored at -20°C until transesterification.

Transesterification of total lipid extracts was performed by adding 1 mL 2% (wt/vol) sulfuric acid in methanol to each sample. Samples were then incubated at 60°C for 2 hours. After that the samples were partially evaporated, and the fatty acid methyl esters (FAME) were extracted by adding 1 mL hexane and vortexing for 10 min. 800 $\mu$L of this hexane was then transferred into glass vials for GC analysis.

GC analysis of FAMEs was performed with a Bruker 450-GC instrument equipped with a flame-ionization detector and a capillary column HP-INNOWAX (30 m $\times$ 0.25 mm). The GC oven conditions were as follows: 150°C (1 min), a 10 min ramp to 230°C, hold at 230°C for 2 min. The split ratio was 10:1. Fatty acids were identified and quantified by comparison with commercial FAME standards normalized to methyl tridecanoate (C13:0). Total lipid content was calculated as the sum of total fatty acid contents for five FAMEs: methyl palmitate (C16:0), methyl palmitoleate (C16:1), methyl stearate (C18:0), methyl oleate (C18:1), methyl linoleate (C18:2) (Sigma-Aldrich). The addition of tridecanoic acid to the chloroform-methanol extraction fluid was used as the internal standard, which was carried through the entire analysis procedure and transesterified into its methyl ester.

## 9.3 Results & Discussion

### 9.3.1 Selection of Acetate Salt

For proof of concept experimentation, it was first necessary to establish the ideal form of acetate for growth. Strain MTYL065 was grown in shake flask experiments with 20 g/L of various acetate salts: sodium acetate, ammonium acetate, potassium acetate. After culturing for 200 hrs, it was found that growth on sodium acetate was significantly better than that of ammonium acetate or potassium acetate, achieving an OD of 10 after 120 hours. The variation may rest in the ability for *Y. lipolytica* to adjust to elevated concentrations of specific ions. As *Y. lipolytica* has been isolated from marine environments, it can tolerate sea water concentrations of sodium chloride (approx. 35 g/L). Salt tolerance is influenced by the presence and activity of salt pumps and *Y. lipolytica* is well known to have strong $Na^+$ pumps as well as an $H^+$-ATPase coupled $Na^+/H^+$ antiporter (Andreishcheva et al., 1999). The necessary sodium pumps help maintain intracellular sodium concentrations in the presence of osmotic stress. It is possible that fewer transport mechanisms are available for ammonium and potassium homeostasis. Additionally, ammonium acetate is possibly an unattractive substrate, since ammonium contributes to the media nitrogen pool. For lipid production, it is necessary to maintain a high C/N ratio - something that cannot be done if ammonium and acetate are fed in stoichiometric quantities.

The growth on sodium acetate proceeded slower than on glucose, achieving an OD of 10 after 5 days, while typical shake flask cultures grown on glucose routinely achieve OD 25 within 4 days (data not shown). There are several possible explanations for the slower growth. In acetate fermentations, the pH actually increases, going from 6 to 9 over the course of a few days, where growth would be strongly inhibited. Furthermore, the elevated sodium concentration may also play a role in lowering growth rates. It is also possible that acetate transport and uptake may be limiting. As acetate metabolism requires increased

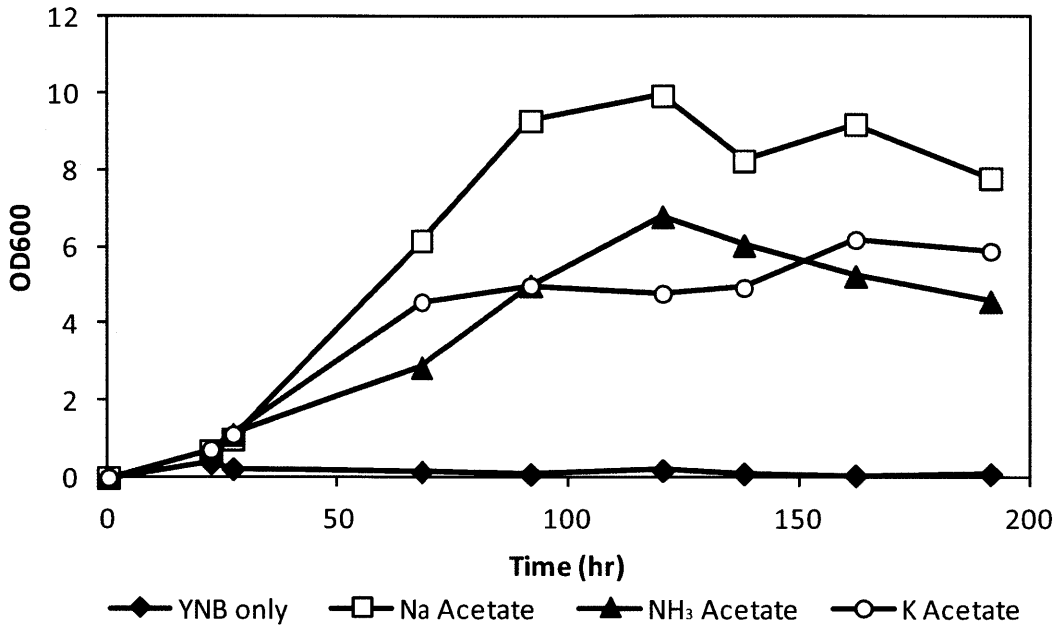


Figure 9.3.1: Growth of *Y. lipolytica* strain MTYL065 on various acetate substrates. 20 g/L of sodium/ammonium/potassium acetate in minimal media was used for 40 mL shake flask studies. A control containing no acetate substrate was included.

flux through the TCA cycle in some organisms (Wendisch et al., 2000), the oxygen-limited conditions could also further limiting growth.

### 9.3.2 Bioreactor Experiment

Once sodium acetate was established as the desired acetate substrate, lipid production performance was tested in a scaled-up 2-L bioreactor. It was found that sodium acetate concentrations above 50 g/L were particularly inhibitory (data not shown), so 50 g/L sodium acetate was charged into the bioreactor, which corresponds to an acetate concentration of approximately 36 g/L. The C/N ratio was adjusted to 100 to match the conditions in glucose fermentation performed in Chapter 6. Since it was speculated that TCA cycle activity might be upregulated, drawing flux towards complete oxidation of acetate, the aeration was kept constant at a low level of 1 vvm. The time course profile of the fermentation is depicted in Figure 9.3.2. After 130 hrs, the acetate was completely consumed, producing 8.9 g/L

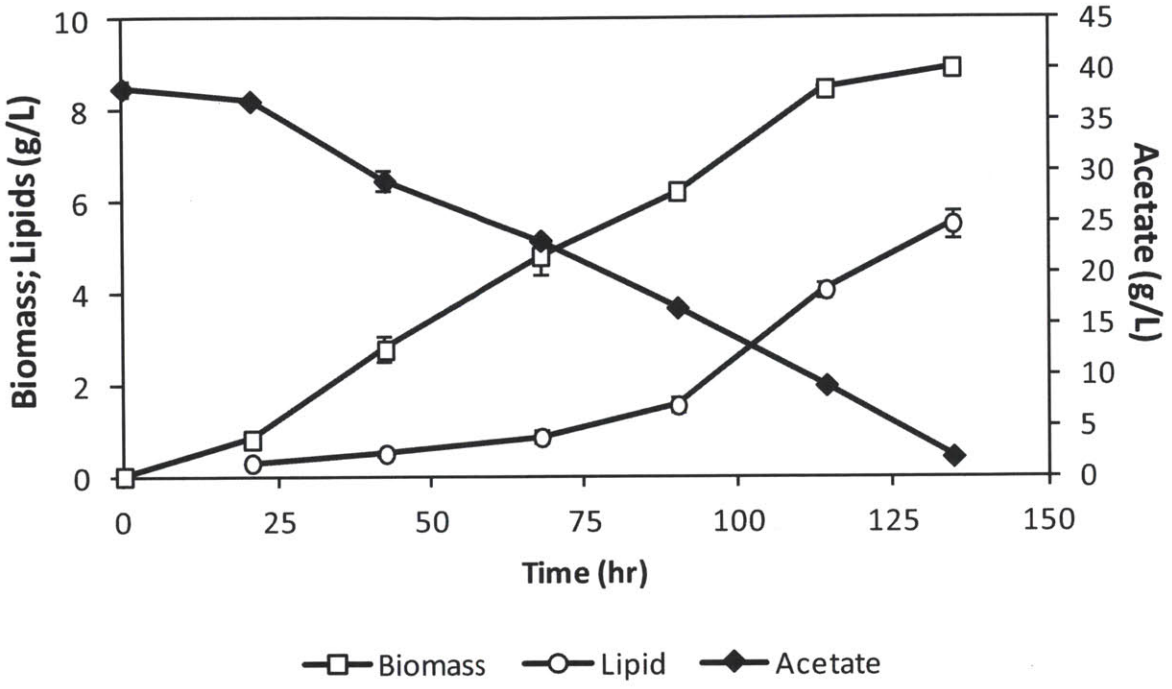


Figure 9.3.2: 2-L bioreactor fermentation of MTYL065 grown on sodium acetate as carbon source.

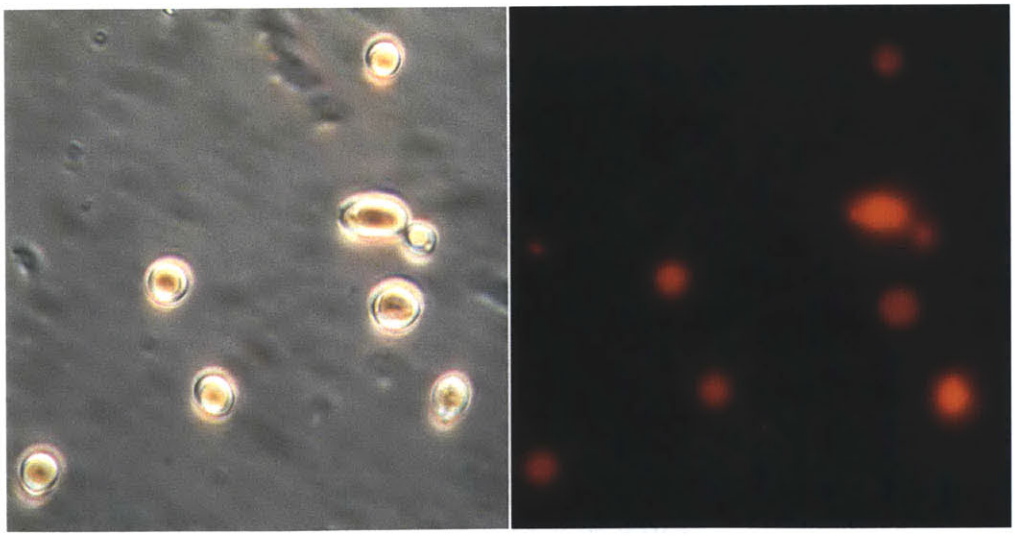

Figure 9.3.3: Microscope image of *Y. lipolytica* after acetate fermentation. Fluorescence microscopy using Nile Red staining highlights lipid fraction (right image, red)

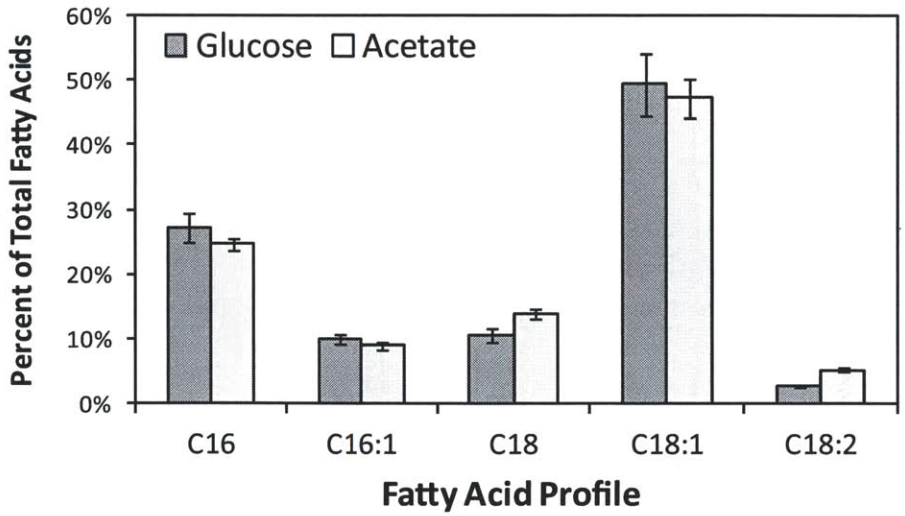


Figure 9.3.4: Comparison of fatty acid profiles of MTYL065 grown on glucose and acetate. Values are given as percent of total fatty acid content for each given fatty acid.

rate through a majority of the fermentation. 5.5 g/L of lipids, or 62% of total biomass, was produced by the end of fermentation, with a lipid production phase beginning after 90 hours. Microscopy of the cells show that the cells are swollen with large lipid bodies; Nile Red staining confirms that these lipid bodies are composed of neutral lipids (Figure 9.3.3). The overall productivity and yield of the process was 0.042 g/L/hr and 0.152 g lipids/g acetate. The fatty acid content, shown in Figure 9.3.4, was indistinguishable from growth on glucose.

When comparing the performance of MTYL065 on acetate vs. glucose, the growth rates and overall yields are significantly lower on acetate. The initial growth rate was found to be roughly one third on acetate as that on glucose. Similarly, the final lipid titer and overall productivity were also one third. Overall, the biomass and lipid yields were also significantly lower on acetate. This is likely because a strong TCA cycle during aerobic fermentation irreversibly consumes acetate, possibly forming a futile metabolic cycle if the energy generated is not needed. This occurred despite the minimal aeration given in the bioreactor. Strong upregulation and flux through the TCA is well observed in other organisms, such as in *Corynebacterium glutamicum*, which increases flux through the TCA cycle 3-fold when

Table 9.1: Comparison of fermentation characteristics in glucose and acetate. C/N ratio of both fermentations were 100, performed in 2-L bioreactors. Glucose reactor initially charged with 90 g/L glucose; Acetate reactor contained 36 g/L acetate (50 g/L sodium acetate).

| | Glucose | Acetate |
|---|---|---|
| Exponential Growth Rate | 0.0982 $hr^{-1}$ | 0.0368 $hr^{-1}$ |
| Overall Biomass Yield | 0.316 g/g | 0.246 g/g |
| Overall Lipid Yield | 0.195 g/g | 0.152 g/g |
| Overall Productivity | 0.147 g/L/hr | 0.041 g/L/hr |
| Lipid Production Phase | 70 - 120 hrs | 90 - 130 hrs |
| Maximum Lipid Yield | 0.249 g/g | 0.270 g/g |
| Maximum Productivity | 0.178 g/L/hr | 0.103 g/L/hr |
| Final Lipid Content | 62 % | 62 % |
| Final Lipid Titer | 17.6 g/L | 5.5 g/L |

grown on acetate instead of glucose(Wendisch et al., 2000). Despite this, the lipid content in both fermentations, given the same C/N ratio, was the same at 62%. This suggests that despite completely different substrates, potential regulation and catabolic pathways, strain MTYL065 activated its lipid synthetic pathway very similarly, diverting and storing similar levels of carbon flux in both cases. Because of the overexpression of ACC and the proximity of acetate to acetyl-coA within the metabolic network, the maximum lipid yield - between 90 and 130 hrs - was calculated to be 0.270 g lipids/g acetate, achieving greater than 90% of theoretical yield. This value is higher than that on glucose, which may incorporate more losses to glycolytic intermediates during glucose catabolism. Despite differences in growth and productivity, strong flux through the lipid synthesis pathway was maintained, demonstrating the robust lipogenic action of *Y. lipolytica* and the engineering improvements in strain MTYL065. In a recent publication on conversion of acetate to lipids in *Y. lipolytica*, a two-stage fermentation was performed where cells were initially grown on glucose but allowed to accumulate lipids with acetate as the fed-batch substrate, controlling the C/N ratio to promote lipid production in the second step only (Fontanille et al., 2012). They report a final lipid content of 40.7% and a yield of 0.13 g lipid/g acetate. Our work demonstrates improved lipid accumulation and lipid yield even with acetate as the sole carbon source. However, their fed-batch system allows much greater biomass production with a

4-fold greater productivity at 0.16 g/L/hr, demonstrating the strength and effectiveness of multi-substrate fermentation systems to build biomass prior to lipid production. Nonetheless our MTYL065 strain performs quite favorably even without process optimization and the use of secondary substrates.

## 9.4 Conclusions

The production of biofuels from electricity holds much promise in providing a land-independent means of producing liquid biofuel. However development of a single organism that can perform this task remains a significant challenge. The use of a two-step reactor system for the conversion of electricity into biofuels with acetate as the intermediate decouples inherent limitations and allows more expedient development of electrofuel technologies. Furthermore, the results of this work demonstrate the robust nature of the metabolically engineered *Y. lipolytica* strain MTYL065. While growth rates and productivities were found to be lower than on glucose, lipid accumulation and maximum lipid yield reached favorable levels. However, much engineering and optimization will still be necessary for a robust process capable of efficiently converting acetate into lipids. For example, as noted earlier, sodium acetate tolerance puts an upper ceiling in a batch fermentation scheme. In future schemes, it may be possible to instead control pH by adding acetic acid rather than the hydrochloric acid normally used, as pH increases with acetate consumption. This could allow for a dynamically controlled fed-batch process where the acetate concentration is kept relatively constant and the culture is allowed to consume acetate as fast as possible. Further improvements in improving growth rate and minimizing excessive carbon loss through TCA cycle oxidation will be needed. However, these results show promise in providing a successful platform for development of the next generation of biofuels.

# References


Andreishcheva, E., Isakova, E., Sidorov, N., Abramova, N., Ushakova, N., Shaposhnikov, G., Soares, M., and Zvyagilskaya, R. (1999) Adaptation to salt stress in a salt-tolerant strain of the yeast *Yarrowia lipolytica. Biochemistry (Mosc) 64*, 1061–1067.

Barth, G., and Gaillardin, C. (1997) Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica. FEMS Microbiol. Rev. 19*, 219–237.

Biello, D. (2011) Let's Go for It. *Scientific American 304*, 30–30.

Folch, J., Lees, M., and Sloane-Stanley, G. H. (1957) A simple method for the isolation and purification of total lipids from animal tissues. *The Journal of Biological Chemistry 226*, 497–509.

Fontanille, P., Kumar, V., Christophe, G., Nouaille, R., and Larroche, C. (2012) Bioconversion of volatile fatty acids into lipids by the oleaginous yeast *Yarrowia lipolytica. Bioresource Technology 114*, 443–449.

Jacobson, M. Z. (2009) Review of solutions to global warming, air pollution, and energy security. *Energy & Environmental Science 2*, 148–173.

Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G., and Nicaud, J. (2001) Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica. Yeast 18*, 97–113.

de la Peña Mattozzi, M., Kang, Y., and Keasling, J. D. In *Handbook of Hydrocarbon and Lipid Microbiology*; Timmis, K. N., Ed.; Springer Berlin Heidelberg, 2010; pp 1649 – 1660.

Stephanopoulos, G. (2007) Challenges in engineering microbes for biofuels production. *Science 315*, 801.

Wendisch, V. F., de Graaf, A. A., Sahm, H., and Eikmanns, B. J. (2000) Quantitative determination of metabolic fluxes during coutilization of two carbon sources: comparative analyses with Corynebacterium glutamicum during growth on acetate and/or glucose. *J Bacteriol 182*, 3088–3096.

# Chapter 10

# Conclusions and Recommendations

As discussed throughout this work, a central goal of thesis has been to explore the utility of metabolic engineering for developing biofuel production platforms. We utilized the yeast *Y. lipolytica* as the host organism and platform for metabolic engineering in the context of yeast oil production. This model organism was selected for its oleaginous nature and robust native lipid production capabilities, as well as the availability of basic genetic tools for metabolic engineering.

## 10.1 Summary of work

Towards this end, we began by establishing additional genetic tools for metabolic engineering in *Y. lipolytica.* In Chapter 4, we identified key features and genetic elements of strong expression found in the promoter and sequence of translation elongation factor 1-$\alpha$, a strong constitutively expressed gene within *Y. lipolytica.* From this we constructed an expression vector utilizing these features, most notably a 5' spliceosomal intron. Through reporter enzyme and RNA analysis, we characterized the promoter performance, showing 12-fold increased expression over the intronless TEF promoter, and 5-fold increased expression over commercial promoter hp4d. These results demonstrated the strong effect of introns on gene expression and how they can be used in expression cassettes to achieve high expresion. Furthermore, we constructed a complementary plasmid based on the vector pACYC-DUET1 to allow for expression of multiple genes within one organism. This plasmid was verified for integration and proper expression and was later used in the combinatorial studies of lipid synthesis. To further understand and characterize the role of spliceosomal introns in yeast, Chapter 5 investigated the functional distribution of introns in a dozen yeast species, identifying strong species-specific enrichment in particular functional pathways.

In Chapter 6 and 7, we utilized the tools established to evaluate rational gene targets in the lipid synthesis pathway. Four targets - ACC, D9, ACL, DGA - were examined as both solitary and combinatorial expression constructs to determine lipid accumulation, lipid

productivity, and lipid yield. It was found that the strain MTYL065, coexpressing ACC and DGA, exhibited the strongest performance. In a scaled-up bioreactor experiment, the strain was able to achieve 62% lipid content at 60% of the theoretical substrate yield. The expression of all four gene targets also showed strong lipid accumulation, with 77% lipid content at 71% theoretical yield. However, comparison of the strain performance highlighted the fact that the increased yield was achieved at the expense of growth. These results demonstrated the efficacy of engineering lipid synthesis in oleaginous microorganisms and synergistic effects that can be obtain through their coexpression.

In Chapter 8, we began to explore how metabolic engineering could be used to extend substrate utilization to alternative substrates. For xylose, a prevalent substrate in cellulosic feedstocks, the native properties of *Y. lipolytica* were examined to determine a functional xylulokinase, but nonfunctional xylose reductase and xylitol dehydrogenase. Subsequent expression of the redox XYL12 pathway from *Scheffersomyces stipitis* and adaptation consisting of serial dilution in minimal media lead to the successful utilization of xylose in *Y. lipolytica*. Modeling of the growth kinetics identified a peculiar, constitutive uptake rate that differed from the substrate-sensitive uptake of glucose. Through the use of cofermentation, growth and productivity on xylose was improved dramatically, with glycerol and xylose identified as a favorable combination for xylose cofermentation. In a scaled-up bioreactor, xylose-to-lipids conversion was successfully achieved, albeit at a low yield.

The alternative substrate acetate, while natively consumed by *Y. lipolytica*, represents an attractive component in a process for non-photosynthetic conversion of carbon dioxide to lipids, with acetate as the intermediate. As a proof of concept for the usefulness of *Y. lipolytica* as a lipid-producing platform in this process, the lipid production performance of our strongest performing strain, MTYL065, was evaluated. The results showed high lipid content and yield, albeit at a slower growth rate than on glucose.

## 10.2 The importance for balance

A common theme developed throughout the course of this work: biological systems require a finely tuned balance to achieve optimal performance. Having too much or too little expression can drastically affect the overall response of the organism.

The functional distribution of introns revealed a fascinating evolutionary journey for yeast as they balanced the benefits of expression-enhancing introns with rapid growth in competitive populations, ultimately leading to the loss of a majority of introns save for those in specific functional pathways.

In the metabolic engineering of lipid biosynthesis, by tipping the balance between lipid storage and mobilization, the DGA gene was found to have a great impact on lipid accumulation, leading to dramatic increases in lipid production. Additionally, the synergistic response of coexpression of ACC and DGA demonstrated the benefits of balancing flux through lipid synthesis by establishing a push and pull dynamic around the feedback inhibiting acyl-coA intermediate. Furthermore, large scale bioreactors comparing the strain containing ACC+DGA with the strain containing all four gene targets highlighted the delicate balance between lipid production and growth, as the latter strain exhibited increased flux through the pathway but at the expense of cell growth. This balance between growth and production has long been a central theme - and obstacle - in metabolic engineering and strain development. The carbon to nitrogen ratio was also observed to have strong effects on fermentation, necessitating careful balance as an important process control factor for lipid production: low values led to suppressed lipid production while high values led to citrate accumulation.

Finally, metabolic engineering of alternative substrate utilization - on both xylose and acetate - emphasized the balance the organism performs between lipid production and respiration, as an overactive respiration response may result in poor yields. With xylose metabolism, the muted metabolic response of the cell to this substrate led to an imbalanced catabolism with undesired loss of substrate to over-respiration.

## 10.3 Recommendations and future work

While this work established tools for engineering *Y. lipolytica*, evaluated manipulations that dramatically improved lipid overproduction, and extended utilization towards alternative substrates, there are still a number of interesting opportunities for further investigation.

Within lipid biosynthesis engineering, a number of other targets could be further investigated. While strong driving force is provided towards lipid production with the genes already examined, further improvements to increase the capacity or maximum rate of lipid productivity would be useful. Identifying subsequent rate-limiting steps and bottlenecks can further improve overall flux towards this end. Additionally, investigating more distal targets, particularly those that may exhibit regulatory function, appears to be an attractive approach. One such target may be the enzyme AMP Kinase, which is a global energy regulator and an inhibitor of the lipid biosynthesis pathway. Mutation or deletion of this enzyme may lead to unrestricted lipid production, perhaps unlocking a phenotype of constitutive lipid production. Other non-rational approaches may also be beneficial, such as transcriptional machinery engineering or inverse metabolic engineering. Other improvements, such as lipid secretion, represent very interesting opportunities to dramatically change the economics and engineering dynamics of the process. However, at this point the mechanism for secretion of triglycerides and whole lipid bodies, which exists in higher eukaryotes, is poorly understood. Another more popular research area is the production of fatty acid derivatives, such as direct FAME production. This removes the need for downstream transesterification and could feasibly provide a drop-in fuel product. However, in *Y. lipolytica* production of free fatty acids is potentially limited by its highly active fatty acid degradation and sequestration pathways.

For xylose metabolism, much work remains available. While the constitutive and constant uptake rate of xylose observed potentially has some advantages, the relatively low rate leaves room for improvement. Properly identifying and alleviating the bottlenecks to xylose utilization will be critical for more robust utilization. This could be through engineering pentose transport, further upregulation of the PPP, or even engineering of cellular regula-

tion. The work with cofermentations demonstrates a helpful first step in xylose utilization in *Y. lipolytica*. The next logical step for cofermentation is the use of *Y. lipolytica* in cellulosic hydrolysate, which contains a mixture of sugars, predominantly glucose and xylose. Testing the utilization and performance of *Y. lipolytica* in this relatively undefined media will demonstrate its future potential in cellulosic biodiesel production, as cellulosic hydrolysate itself has its own challenges, containing numerous inhibitory compounds that often lower productivity. Furthermore, the low yields on xylose during the bioreactor cofermentation highlight the need for further understanding and engineering of the overall metabolism of *Y. lipolytica* to optimally channel carbon towards lipid production. More comprehensive coupling of lipid synthesis engineering and xylose metabolism engineering could be a step in this direction.

Finally with acetate utilization, while yields and performance on acetate are very promising, there is still much to be done in terms of process development for robust production of lipids from acetate. The introduction of novel fed-batch schemes, such as acetic acid pH control, promises to greatly improve control and robustness of the fermentation. Additionally, identifying possible bottlenecks in acetate fermentation, whether it is oxygen-limitation or acetate uptake, will be necessary to fully understand and characterize the system.

The field of metabolic engineering continues to mature as it begins to transition into commodity products such as biofuels. It stands at the interface of traditional chemical engineering and state-of-the-art biotechnology, in the geopolitical context of alternative energy development. It will be critical to utilize the existing knowledge base and experience to properly develop and engineer platforms that can successfully be used at the industrial scale. Through metabolic engineering of the oleaginous yeast *Y. lipolytica*, we have shown that through utilization of this knowledge base to make selective perturbations to cellular metabolism, we can achieve robust overproduction of lipids from a variety of substrates. These results demonstrate the great promise of *Y. lipolytica* as a model organism both for studying lipid overproduction and metabolic engineering for biofuels development.

# Appendix A

# Intron Bioinformatics Code

Table A.1: Intron Bioinformatics Supporting Files

| File name | Comments |
|---|---|
| yeast_introns.pl | Perl script used to retrieve list of genes with introns |
| PID Table | Table of KEGG pathway PID values and their corresponding pathways |
| genes.txt | Text file containing total number of genes for each species in each pathway |
| introns.txt | Text file containing total number of introns for each species in each pathway |
| pvaluecalculation.m | MATLAB code to calculate enrichment p-value from intron list |
| plotenrichmentscore.m | Generates enrichment scores based on pvalues and graphically plots data |
| h2fconvert.m | Converts sample population statistics into a contingency table |
| FisherExactTest.m | Performs Fisher Exact Test based on contingency table input |

## A.1 Introduction

Code was written in order to retrieve, analyze and display intron functional distributions. First, the list of pathway genes for each organism is retrieved from the KEGG database. A PERL script (yeast_introns.pl) is then run which checks NCBI (http://www.ncbi.nlm.nih.gov/) for the presence of introns for each target gene. The resulting list of hits is then reorganized and tabulated by using a gene-to-KEGG pathway mapping file (PID Table) These tabulated results are then organized into two tables for all organisms: a table of the number of KEGG genes per category, and the number of introns identified per category (genes.txt, introns.txt). MATLAB scripts are run to calculate overenrichment from the categorized hit table (pvalue-calculation.m). Finally the statistical data is compiled and displayed into a graphical chart (plotenrichmentscore.m). Helper files are used in performing the statistical Fisher exact test (h2fconvert.m, FisherExactTest.m). As of July 1, 2011, the gene listings for each organism from KEGG (on their FTP site) are accessible only to paid subscribers. These results were obtained prior to this policy change.

## A.2 yeast_introns.pl

```
open(INPUT, "<xxx_url_list.txt");
my @text = <INPUT>;

open(OUTPUT, ">xxx_introns_results.txt");

print OUTPUT "gene,intron";
print OUTPUT "\n";

my $i = 0;
my $intron;

while ($i < 99999) {
    my $url = $text[$i];

    chomp ($url);
    chomp ($url);
    print $url;

    use LWP::Simple;
    my $content = get $url;

    $intron = 0;

    if ($content =~  m/ exons,/) s{
        if ($content =~ m/ 1 exons/) {

        }
        else {
            $intron = 1;
        }
    }
```

```
    print $intron;
    print "\n";

    print OUTPUT "$url";
    print OUTPUT ",";
    print OUTPUT "$intron";
    print OUTPUT "\n";
    $i = $i + 1;
}

print $intron;
```

## A.3 KEGG PID Table

Table A.2: KEGG PID Table (PID 10 - 240)

| PID | Pathway | Pathway Category | Heirarchy | Module |
|---|---|---|---|---|
| 10 | Glycolysis / Gluconeogenesis | Carbohydrate Metabolism | Metabolism | Central carbohydrate metabolism |
| 20 | Citrate cycle (TCA cycle) | Carbohydrate Metabolism | Metabolism | Central carbohydrate metabolism |
| 30 | Pentose phosphate pathway | Carbohydrate Metabolism | Metabolism | Central carbohydrate metabolism |
| 40 | Pentose and glucuronate interconversions | Carbohydrate Metabolism | Metabolism | Other carbohydrate metabolism |
| 51 | Fructose and mannose metabolism | Carbohydrate Metabolism | Metabolism | Nucleotide sugar metabolism |
| 52 | Galactose metabolism | Carbohydrate Metabolism | Metabolism | Nucleotide sugar metabolism |
| 53 | Ascorbate and aldarate metabolism | Carbohydrate Metabolism | Metabolism | Cofactors and vitamins |
| 61 | Fatty acid biosynthesis | Lipid Metabolism | Metabolism | Fatty acid metabolism |
| 62 | Fatty acid elongation in mitochondria | Lipid Metabolism | Metabolism | Fatty acid metabolism |
| 71 | Fatty acid metabolism | Lipid Metabolism | Metabolism | Fatty acid metabolism |
| 72 | Synthesis and degradation of ketone bodies | Lipid Metabolism | Metabolism | Fatty acid metabolism |
| 100 | Steroid biosynthesis | Lipid Metabolism | Metabolism | Sterol metabolism |
| 130 | Ubiquinone and other terpenoid-quinone biosynthesis | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 190 | Oxidative phosphorylation | Energy Metabolism | Metabolism | ATP synthesis |
| 230 | Purine metabolism | Nucleotide Metabolism | Metabolism | Purine metabolism |
| 232 | Caffeine metabolism | Biosynthesis of Other Secondary Metabolites | Metabolism | Purine metabolism |
| 240 | Pyrimidine metabolism | Nucleotide Metabolism | Metabolism | Pyrimidine metabolism |

Table A.3: KEGG PID Table (PID 250 - 430)

| PID | Pathway | Pathway Category | Heirarchy | Module |
|---|---|---|---|---|
| 250 | Alanine, aspartate and glutamate metabolism | Amino Acid Metabolism | Metabolism | Alanine, aspartate and glutamate metabolism |
| 260 | Glycine, serine and threonine metabolism | Amino Acid Metabolism | Metabolism | Glycine, serine and threonine metabolism |
| 270 | Cysteine and methionine metabolism | Amino Acid Metabolism | Metabolism | Cysteine and methionine metabolism |
| 280 | Valine, leucine and isoleucine degradation | Amino Acid Metabolism | Metabolism | Branched-chain amino acid metabolism |
| 290 | Valine, leucine and isoleucine biosynthesis | Amino Acid Metabolism | Metabolism | Branched-chain amino acid metabolism |
| 300 | Lysine biosynthesis | Amino Acid Metabolism | Metabolism | Lysine metabolism |
| 310 | Lysine degradation | Amino Acid Metabolism | Metabolism | Lysine metabolism |
| 311 | Penicillin and cephalosporin biosynthesis | Biosynthesis of Other Secondary Metabolites | Metabolism | Lysine metabolism |
| 330 | Arginine and proline metabolism | Amino Acid Metabolism | Metabolism | Arginine and proline metabolism |
| 340 | Histidine metabolism | Amino Acid Metabolism | Metabolism | Histidine metabolism |
| 350 | Tyrosine metabolism | Amino Acid Metabolism | Metabolism | Aromatic amino acid metabolism |
| 360 | Phenylalanine metabolism | Amino Acid Metabolism | Metabolism | Aromatic amino acid metabolism |
| 380 | Tryptophan metabolism | Amino Acid Metabolism | Metabolism | Aromatic amino acid metabolism |
| 400 | Phenylalanine, tyrosine and tryptophan biosynthesis | Amino Acid Metabolism | Metabolism | Aromatic amino acid metabolism |
| 410 | beta-Alanine metabolism | Metabolism of Other Amino Acids | Metabolism | Other amino acid metabolism |
| 430 | Taurine and hypotaurine metabolism | Metabolism of Other Amino Acids | Metabolism | Other amino acid metabolism |

Table A.4: KEGG PID Table (PID 450 - 563)

| PID | Pathway | Pathway Category | Heirarchy | Module |
|---|---|---|---|---|
| 450 | Selenoamino acid metabolism | Metabolism of Other Amino Acids | Metabolism | Other amino acid metabolism |
| 460 | Cyanoamino acid metabolism | Metabolism of Other Amino Acids | Metabolism | Other amino acid metabolism |
| 472 | D-Arginine and D-ornithine metabolism | Metabolism of Other Amino Acids | Metabolism | Other amino acid metabolism |
| 480 | Glutathione metabolism | Metabolism of Other Amino Acids | Metabolism | Cofactors and vitamins |
| 500 | Starch and sucrose metabolism | Carbohydrate Metabolism | Metabolism | Nucleotide sugar metabolism |
| 510 | N-Glycan biosynthesis | Glycan Biosynthesis and Metabolism | Metabolism | Glycoprotein metabolism |
| 511 | Other glycan degradation | Glycan Biosynthesis and Metabolism | Metabolism | Glycoprotein metabolism |
| 513 | High-mannose type N-glycan biosynthesis | Glycan Biosynthesis and Metabolism | Metabolism | Glycoprotein metabolism |
| 514 | O-Mannosyl glycan biosynthesis | Glycan Biosynthesis and Metabolism | Metabolism | Glycoprotein metabolism |
| 520 | Amino sugar and nucleotide sugar metabolism | Carbohydrate Metabolism | Metabolism | Nucleotide sugar metabolism |
| 531 | Glycosaminoglycan degradation | Glycan Biosynthesis and Metabolism | Metabolism | Glycosaminoglycan metabolism |
| 561 | Glycerolipid metabolism | Lipid Metabolism | Metabolism | Glycerolipid metabolism |
| 562 | Inositol phosphate metabolism | Carbohydrate Metabolism | Metabolism | Glycerophospholipid metabolism |
| 563 | Glycosylphosphatidylinositol (GPI)-anchor biosynthesis | Glycan Biosynthesis and Metabolism | Metabolism | GPI-anchor metabolism |

Table A.5: KEGG PID Table (PID 564 - 750)

| PID | Pathway | Pathway Category | Heirarchy | Module |
|---|---|---|---|---|
| 564 | Glycerophospholipid metabolism | Lipid Metabolism | Metabolism | Glycerophospholipid metabolism |
| 565 | Ether lipid metabolism | Lipid Metabolism | Metabolism | Glycerophospholipid metabolism |
| 590 | Arachidonic acid metabolism | Lipid Metabolism | Metabolism | Fatty acid metabolism |
| 592 | alpha-Linolenic acid metabolism | Lipid Metabolism | Metabolism | Fatty acid metabolism |
| 600 | Sphingolipid metabolism | Lipid Metabolism | Metabolism | Sphingolipid metabolism |
| 603 | Glycosphingolipid biosynthesis - globo series | Glycan Biosynthesis and Metabolism | Metabolism | Glycolipid metabolism |
| 620 | Pyruvate metabolism | Carbohydrate Metabolism | Metabolism | Central carbohydrate metabolism |
| 630 | Glyoxylate and dicarboxylate metabolism | Carbohydrate Metabolism | Metabolism | Other carbohydrate metabolism |
| 640 | Propanoate metabolism | Carbohydrate Metabolism | Metabolism | Other carbohydrate metabolism |
| 650 | Butanoate metabolism | Carbohydrate Metabolism | Metabolism | Other carbohydrate metabolism |
| 670 | One carbon pool by folate | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 680 | Methane metabolism | Energy Metabolism | Metabolism | Other energy metabolism |
| 730 | Thiamine metabolism | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 740 | Riboflavin metabolism | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 750 | Vitamin B6 metabolism | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |

Table A.6: KEGG PID Table (PID 760 - 1053)

| PID | Pathway | Pathway Category | Heirarchy | Module |
|---|---|---|---|---|
| 760 | Nicotinate and nicotinamide metabolism | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 770 | Pantothenate and CoA biosynthesis | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 780 | Biotin metabolism | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 785 | Lipoic acid metabolism | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 790 | Folate biosynthesis | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 860 | Porphyrin and chlorophyll metabolism | Metabolism of Cofactors and Vitamins | Metabolism | Cofactors and vitamins |
| 900 | Terpenoid backbone biosynthesis | Biosynthesis of Polyketides and Terpenoids | Metabolism | Terpenoids |
| 903 | Limonene and pinene degradation | Biosynthesis of Polyketides and Terpenoids | Metabolism | Terpenoids |
| 910 | Nitrogen metabolism | Energy Metabolism | Metabolism | Other energy metabolism |
| 920 | Sulfur metabolism | Energy Metabolism | Metabolism | Other energy metabolism |
| 970 | Aminoacyl-tRNA biosynthesis | Translation | Genetic Information Processing | Translation |
| 980 | Metabolism of xenobiotics by cytochrome P450 | Xenobiotics Biodegradation and Metabolism | Metabolism | Xenobiotics Biodegradation and Metabolism |
| 1040 | Biosynthesis of unsaturated fatty acids | Lipid Metabolism | Metabolism | Fatty acid metabolism |
| 1053 | Biosynthesis of siderophore group nonribosomal peptides | Biosynthesis of Polyketides and Terpenoids | Metabolism | Terpenoids |

Table A.7: KEGG PID Table (PID 1100 - 3410)

| PID | Pathway | Pathway Category | Heirarchy | Module |
|---|---|---|---|---|
| 1100 | Metabolic Pathways | Metabolic Pathway (GLOBAL) | Global | Global |
| 1110 | Biosynthesis of secondary metabolites | Metabolic Pathway (GLOBAL) | Global | Global |
| 2010 | ABC transporters | Membrane Transport | Environmental Information Processing | Membrane Transport |
| 3010 | Ribosome | Translation | Genetic Information Processing | Translation |
| 3018 | RNA degradation | Folding, Sorting and Degradation | Genetic Information Processing | Folding |
| 3020 | RNA polymerase | Transcription | Genetic Information Processing | Transcription |
| 3022 | Basal transcription factors | Transcription | Genetic Information Processing | Transcription |
| 3030 | DNA replication | Replication and Repair | Genetic Information Processing | Replication |
| 3040 | Spliceosome | Transcription | Genetic Information Processing | Transcription |
| 3050 | Proteasome | Folding, Sorting and Degradation | Genetic Information Processing | Folding |
| 3060 | Protein export | Membrane Transport | Environmental Information Processing | Membrane Transport |
| 3410 | Base excision repair | Replication and Repair | Genetic Information Processing | Repair |

Table A.8: KEGG PID Table (PID 3420 - 4650)

| PID | Pathway | Pathway Category | Heirarchy | Module |
|---|---|---|---|---|
| 3420 | Nucleotide excision repair | Replication and Repair | Genetic Information Processing | Repair |
| 3430 | Mismatch repair | Replication and Repair | Genetic Information Processing | Repair |
| 3440 | Homologous recombination | Replication and Repair | Genetic Information Processing | Repair |
| 3450 | Non-homologous end-joining | Replication and Repair | Genetic Information Processing | Repair |
| 4011 | MAPK signaling pathway - yeast | Signal Transduction | Environmental Information Processing | Signal Transduction |
| 4070 | Phosphatidylinositol signaling system | Signal Transduction | Environmental Information Processing | Signal Transduction |
| 4111 | Cell cycle - yeast | Cell Growth and Death | Cellular Processes | Replication |
| 4120 | Ubiquitin mediated proteolysis | Folding, Sorting and Degradation | Genetic Information Processing | Folding |
| 4130 | SNARE interactions in vesicular transport | Folding, Sorting and Degradation | Genetic Information Processing | Folding |
| 4140 | Regulation of autophagy | Transport and Catabolism | Cellular Processes | Transport and Catabolism |
| 4144 | Endocytosis | Transport and Catabolism | Cellular Processes | Transport and Catabolism |
| 4145 | Phagosome | Transport and Catabolism | Cellular Processes | Transport and Catabolism |
| 4146 | Peroxisome | Transport and Catabolism | Cellular Processes | Transport and Catabolism |
| 4650 | Natural killer cell mediated cytotoxicity | Immune System | Organismal Systems | Immune System |

## A.4 pvaluecalculation.m

```
% Hypergeometric Distribution Test for Intron Over-Representation
% Calculates the p-value of an observed intron density with respect to
% the overall intron density
% inputs are two files: introns.txt and genes.txt, which contain the intron
% counts and gene counts for each species in a tab-delimited text file
% 2/3/2011
function pvaluecalculation()
clear all;
clc
organisms = {'CTP'  'Candida tropicalis';
'CGR'   'Candida glabrata';
'KLA'   'Kluyveromyces lactis';
'SCE'   'Saccharomyces cereivisiae';
'AGO'   'Ashbya gossypii';
'DHA'   'Debaryomyces hansenii';
'PPA'   'Pichia pastoris';
'PIC'   'Pichia stipitis';
'YLI'   'Yarrowia lipolytica';
'SPO'   'Schizosaccharomyces pombe';
'ZRO'   'Zygosaccharomyces rouxii';
'NCR'   'Neurospora crassa'
};
pathways = {'Carbohydrate Metabolism';
'Energy Metabolism';
'Lipid Metabolism';
'Amino Acid Metabolism';
'Metabolism of Other Amino Acids';
'Metabolism of Cofactors and Vitamins';
'Transcription';
'Translation';
'Folding, Sorting and Degradation';
```

```
'Replication and Repair';
'Transport and Catabolism';
'Cell Growth and Death'};
% Load intron and gene counts for all organisms
fid = fopen('introns.txt');
introns = fscanf(fid, '%g', [length(organisms) inf]);
fclose(fid);

fid = fopen('genes.txt');
genes = fscanf(fid, '%g', [length(organisms) inf]);
fclose(fid);
% variable(organism index, pathway index)

%for each organism:
for i=1:length(organisms),
    disp(organisms(i,2));
    total_introns = sum(introns(i,:));
    total_genes = sum(genes(i,:));
    for k=1:length(pathways),
        [a b c d] = h2fconvert(introns(i,k),genes(i,k),total_introns,total_genes);
        p = FisherExactTest([a b; c d]);
        % Because there is an error in the Fisher Test code (sometimes
        % the negative and positive tail p-values will be switched
        % just determine whether sampling is higher or lower than
        % population average, and take the lower/higher value, respectively
        if(introns(i,k)/genes(i,k) > total_introns/total_genes)
            pDESIRED = min(p(1:2));
        else
            pDESIRED = max(p(1:2));
        end
       pvalue(i,k)= pDESIRED;
    end
end
```

```
introns'
pval = pvalue'

fid = fopen('pvalue.dat','w');
fwrite(fid, pvalue', 'double');
fclose(fid);
surf(log(pvalue));

score = -log(pvalue)'
```

## A.5 plotenrichmentscore.m

```
% PlotEnrichment Score
% Reads in the p-values of each category-species fisher exact test
% and rearranges them for presentability purposes. also presents the
% results as enrichment scores -log(p-value)
function plotenrichmentscore()
clc; clear all; clf; close all;
organisms = {'CTP'  'Candida tropicalis';
'CGR'   'Candida glabrata';
'KLA'   'Kluyveromyces lactis';
'SCE'   'Saccharomyces cereivisiae';
'AGO'   'Ashbya gossypii';
'DHA'   'Debaryomyces hansenii';
'PPA'   'Pichia pastoris';
'PIC'   'Pichia stipitis';
'YLI'   'Yarrowia lipolytica';
'SPO'   'Schizosaccharomyces pombe';
'ZRO'   'Zygosaccharomyces rouxii';
'NCR'   'Neurospora crassa'
};
pathways = {'Carbohydrate Metabolism';
'Energy Metabolism';
'Lipid Metabolism';
'Amino Acid Metabolism';
'Metabolism of Other Amino Acids';
'Metabolism of Cofactors and Vitamins';
'Transcription';
'Translation';
'Folding, Sorting and Degradation';
'Replication and Repair';
'Transport and Catabolism';
'Cell Growth and Death'};
```

```
% Phylogenic rank of organisms ("PHYL.RANK")
% rank of each organism rank = [8 7 2 4 1 5 8 6 9 10 11 3];
% rearrange so that the index is that rank ("RE-INDEX")
rank = [4 2 11 3 5 7 1 9 6 8 12 10];

% Load intron and gene counts for all organisms,
% and p-values calculated from pvaluecalculation.m
fid = fopen('introns.txt');
introns = fscanf(fid, '%g', [length(organisms) inf]);
fclose(fid);

fid = fopen('genes.txt');
genes = fscanf(fid, '%g', [length(organisms) inf]);
fclose(fid);

fid = fopen('pvalue.dat');
pvalues = fread(fid, [length(pathways),length(organisms)], 'double');
fclose(fid);
% variable(pathway index, organism index)

pvalues
% sort the organisms in order of phylogeny, using ranking

for i=1:length(organisms)-1,
    z = rank(i);
    for k=1:length(pathways),
        dispvalues(k,i) = pvalues(k,z);
    end
end
%sort the pathways in order of magnitude (so the graph looks better)
pathrank = [6 5 4 3 1 2 12 10 11 9 7 8];
values = abs(log(dispvalues));
for i=1:length(pathways)
```

```
    z = pathrank(i);
    for k=1:length(organisms)-1
        finalvalues(i,k) = values(z,k);
    end
end
% Az -160 Ev 30
orglabel = ['SCE';
    'CGR';
    'ZRO';
    'KLA';
    'AGO';
    'PPA';
    'CTP';
    'YLI';
    'DHA';
    'PIC';
    'NCR';
    'SPO'];
pathlabel = {'Metabolism of Cofactors and Vitamins';
   'Metabolism of Other Amino Acids';
   'Amino Acid Metabolism';
   'Lipid Metabolism';
   'Carbohydrate Metabolism';
   'Energy Metabolism';
   'Cell Growth and Death';
   'Replication and Repair';
   'Transport and Catabolism';
   'Folding, Sorting and Degradation';
   'Transcription';
   'Translation'};

finalvalues
h = bar3(finalvalues,1);
```

```
hh = get(h(3),'parent');
%hh2 = get(h(2),'parent');
set(hh,'xticklabel',orglabel);
set(hh,'yticklabel',pathlabel);
view(-160,30);
grid off;
colormap('white');
zlabel('Enrichment Score');
%ylabel('KEGG Pathway');
%xlabel('Organism');
%title('KEGG-Intron Over-Representation Analysis');
```

## A.6 h2fconvert.m

```
% h2f convert
% Simple function which calculates the elements of a contigency table
% given sample and total sizes, sample and total success rates
function [a, b, c, d] = h2fconvert(sample_succ, sample_size, total_succ, total_size)
b = sample_succ;
a = sample_size - sample_succ;
d = total_succ - sample_succ;
c = total_size - total_succ - sample_size + sample_succ;
```

## A.7 FisherExactTest.m

```
function [ Pvalue ConTable ]=FisherExactTest22( x,y )
% Fisher's Exact Probability Test for contigency table 2*2.
% Explanation: http://mathworld.wolfram.com/FishersExactTest.html
% Also refer to: http://www.mathworks.com/matlabcentral/fileexchange/15434
% Inputs:
% Choice (1) one variable x - 2x2 data matrix
% Choice (2) two variable x, y for binary vector with the same length.(No missing val
% Outputs:
% (1) four p-values organized as:
% Pvalue = [ left tail , Right tail, 2-tails, Mid-p correction for 2-tails ]
% Created by Lowell Guangdi 2010/01/28
if nargin > 1
ConTable = zeros(2);
for p = 1:length(x)
if x( p ) == 0 && y( p ) == 0, ConTable(1,1) = ConTable(1,1) + 1;
elseif x( p ) == 0 && y( p ) == 1 ,ConTable(1,2) = ConTable(1,2) + 1;
elseif x( p ) == 1 && y( p ) == 0 ,ConTable(2,1) = ConTable(2,1) + 1;
else ConTable(2,2) = ConTable(2,2) + 1;
end
end
x = ConTable;
else
ConTable = x;
end
Rs=sum(x,2); Cs=sum(x); N=sum(Rs);
%Rearrange the matrix if necessary.
if ~issorted(Rs),x=flipud(x); Rs=sort(Rs); end
if ~issorted(Cs),x=fliplr(x); Cs=sort(Cs); end
A=0:1:min(Rs(1),Cs(1));
z=[A;Rs(2)-Cs(1)+A;Rs(1)-A;Cs(1)-A;];
np=zeros(1,length(A)); lz=log(z);
```

```
np(1)=sum(gammaln([Rs(2)+1 Cs(2)+1])-gammaln([N+1 z(2)+1]));
f=sum(lz(3:4,1:end-1))-sum(lz(1:2,2:end));
np(2:end)=np(1)+cumsum(f);
np=exp(np); W=x(1)+1;
%now compute the 1-tailed p-values and the 2-tailed p-value
if x(1)<=round(A(end)/2) %choose direction
P=[sum(np(1:W)) sum(np(W:end)) sum(np(np<=np(W)))];
else
P=[sum(np(W:end)) sum(np(1:W)) sum(np(np<=np(W)))];
end
Pvalue = [ P 0.5*np(W)+sum(np(np<np(W))) ];
end
```